

n° 147  
2 microfiches

Joëlle  
CROS épouse LEFER

**T h è s e s   e t   d o c u m e n t s   m i c r o f i c h é s**

Implications phylogénétiques  
des variations de l'ADN chloroplastique  
chez les caféiers  
(genres *Coffea* L. et *Psilanthus* Hook. f.)

**ORSTOM**  
éditions

ACADEMIE DE MONTPELLIER

UNIVERSITE MONTPELLIER II

- SCIENCES ET TECHNIQUES DU LANGUEDOC -

## THESE

présentée à l'Université de Montpellier II Sciences et Techniques du Languedoc  
pour obtenir le diplôme de DOCTORAT

SPECIALITE : BIOLOGIE DES POPULATIONS ET ECOLOGIE  
Formation Doctorale : Evolution et Ecologie  
Ecole Doctorale : Biologie des systèmes intégrés - Agronomie, Environnement

# **Implications phylogénétiques des variations de l'ADN chloroplastique chez les Caféiers (genres *Coffea* L. et *Psilanthus* Hook. f.)**

par

**Joëlle CROS épouse LEFER**

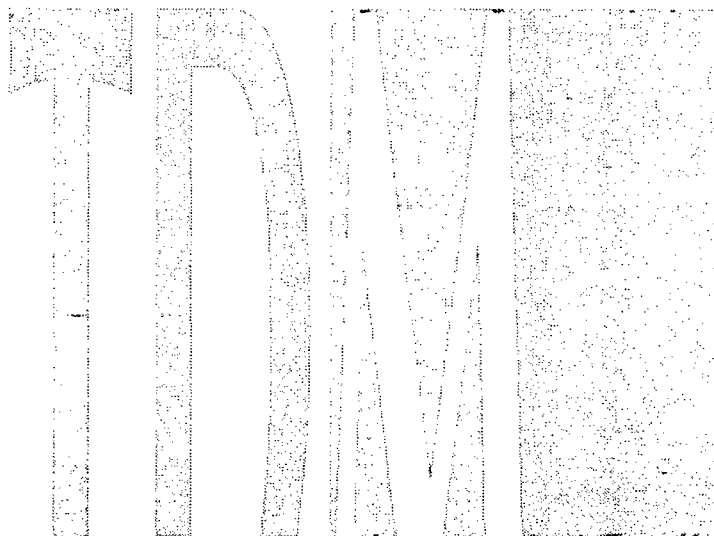
Soutenue le 19 décembre 1994 devant le jury composé de :

MM. Bernard Delay, Directeur de Recherche, Université Montpellier II.  
J.-L. Guillaumet, Directeur de Recherche, ORSTOM Paris.  
Pierre Taberlet, Chargé de Recherche, CNRS Grenoble.  
Philippe Lashermes, Chargé de Recherches, ORSTOM Montpellier.  
André Charrier, Professeur, ENSA Montpellier.

Président  
Rapporteur  
Rapporteur  
Examineur  
Directeur de Thèse

n° 147

2 microfiches



Thèses et documents microfichés



Orstom, l'Institut français de recherche scientifique  
pour le développement en coopération

La loi du 11 mars 1957 n'autorisant, aux termes des alinéas 2 et 3 de l'article 41, d'une part, que les «copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective» et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, «toute représentation ou reproduction intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause, est illicite» (alinéa 1<sup>er</sup> de l'article 40).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles 425 et suivants du Code pénal.

© ORSTOM éditions 1996, Paris

ISSN : 0767-922X

ISBN : 2-7099-1313-5





## Résumé :

Les variations de l'ADN chloroplastique de vingt-sept taxons de caféiers représentant les principales espèces et six taxons indéterminés ont été étudiées. Le génome chloroplastique des caféiers montre globalement un faible polymorphisme. La séquence intergénique *trnL-trnF* apparaît plus polymorphe. L'analyse phylogénétique de ces données moléculaires suggère une explosion radiative largement postérieure à la dislocation du Gondwana, et une organisation en au moins six clades. Cette organisation diffère de la structure en trois groupes "biogéographiques" proposés à la suite des études de diversité agro-morphologique et biochimique. Des hypothèses sont avancées quant à l'origine et la spéciation des caféiers en relation avec l'histoire de l'Afrique et de Madagascar.

## Mots-clés :

ADN chloroplastique, phylogénie, séquence intergénique, Afrique, *Coffea*, RFLP, taille du génome.

## **SOMMAIRE**

<b>INTRODUCTION</b>	<b>4</b>
<b>1. LES CAFEIERS</b>	<b>7</b>
<b>2. L'ADN CHLOROPLASTIQUE POUR LA RECONSTRUCTION DES PHYLOGENIES</b>	<b>31</b>
<b>3. VARIATIONS DE L'ADN CHLOROPLASTIQUE DES CAFEIERS</b>	<b>51</b>
<b>4. VARIATIONS DE REGIONS NON CODANTES DU GENOME CHLOROPLASTIQUE DES CAFEIERS</b>	<b>67</b>
<b>5. PHYLOGENIE DES CAFEIERS A PARTIR DE DONNEES CHLOROPLASTIQUES</b>	<b>90</b>
<b>6. VARIATION DE LA TAILLE DU GENOME NUCLEAIRE DES CAFEIERS ET RELATIONS ENTRE ESPECES</b>	<b>113</b>
<b>7. DISCUSSION CONCLUSION</b>	<b>126</b>
<b>REFERENCES BIBLIOGRAPHIQUES</b>	<b>139</b>
<b>ANNEXES</b>	<b>150</b>

# **INTRODUCTION**

Les caféiers appartiennent à la famille des Rubiaceae. Ils sont originaires d'Afrique et de Madagascar. Le genre *Coffea* sous-genre *Coffea* qui est le plus riche et le plus intéressant du point de vue de ses potentialités agronomiques comprend plus de 80 espèces dont seulement 2 sont cultivées : *Coffea arabica* et *Coffea canephora*. Une troisième espèce, *C. liberica*, a été cultivée au début du siècle avant d'être décimée par la trachéomycose (*Fusarium xylarioides*). Elle représente aujourd'hui moins de 1% de la production mondiale. *C. arabica* représente à elle seule 80% de la production mondiale actuelle. Mais, comme beaucoup d'autres espèces tropicales introduites pour la culture sur d'autres continents entre le XVI<sup>ème</sup> et le XX<sup>ème</sup> siècle, cette espèce possède une base génétique très restreinte qui la rend particulièrement vulnérable aux épidémies. Chez une plante d'importance économique telle que le caféier, l'accroissement de la variabilité génétique et l'apport de nouvelles potentialités (résistances aux maladies et parasites, qualité du produit) est donc un enjeu majeur pour la recherche. Contrairement à l'espèce *C. arabica*, l'ensemble des espèces sauvages de caféiers présente au contraire une diversité génétique importante. De nombreux travaux ont donc été entrepris pour l'étudier. Les éléments recueillis sont multiples mais limités le plus souvent à quelques espèces. Ils concernent principalement des données morphologiques (Anthony, 1992 ; Berthaud *et al.*, 1984 ; Carvalho *et al.*, 1991 ), isoenzymatiques (Anthony, 1992 ; Berthou *et al.*, 1980) et biochimiques (Anthony *et al.*, 1993 ; Rakotomalala *et al.*, 1993 ; Santa Ram *et al.*, 1982). L'analyse de l'ensemble de ces données a permis de proposer une structuration des espèces du sous-genre *Coffea* en trois ensembles biogéographiques (guinéo-congolais, est-africain et malgache). Toutefois, faute d'avoir employé des marqueurs adaptés à l'étude de la phylogénie, la connaissance des relations de parenté entre espèces reste imprécise.

L'objectif de notre étude était d'aborder le problème des relations de parenté entre espèces. Notre travail s'est appuyé sur une approche moléculaire de la phylogénie. Il s'est axé principalement sur l'étude des variations de l'ADN chloroplastique qui présente pour cela plusieurs avantages :

- Tout d'abord les avantages intrinsèques aux études de niveau moléculaire qui sont liés principalement à l'augmentation considérable de la panoplie des caractères disponibles. Elles autorisent notamment l'analyse directe du génome, dans sa totalité (parties codantes et non-codantes) et indépendamment du stade de développement de la plante. Ce type d'étude produit généralement des phylogénies cohérentes et robustes.

- Ensuite les avantages liés au choix de l'ADN chloroplastique. Chez les plantes supérieures l'ADN chloroplastique qui est abondant et conservé constitue un support privilégié pour l'étude des relations de parenté. Il produit des reconstructions phylogénétiques qui comportent généralement peu d'homoplasie.

Pour réaliser cette étude nous avons bénéficié du matériel récolté au cours des différentes missions de prospection en Afrique et à Madagascar. Le matériel végétal a été choisi afin de représenter au mieux la diversité génétique en se basant sur les études antérieures concernant des caractéristiques morphologiques, izoenzymatiques et biochimiques. Des taxons récemment prospectés et sur lesquels nous possédons beaucoup moins d'informations, ont également été inclus dans l'analyse.

Ce mémoire est développé en sept parties :

- La première partie a trait au matériel végétal. Elle consiste en une présentation de la taxonomie des espèces de caféiers et les données concernant leur diversité. Nous abordons également les hypothèses paléogéographiques, paléobotaniques et biogéographiques susceptibles de nous aider dans l'interprétation phylogénétique de la diversité observée.
- La deuxième partie concerne les potentialités de l'approche moléculaire appliquée au génome chloroplastique. Nous présentons le type d'informations produit et les modes d'analyses envisageables.

Ces deux premières parties bibliographiques sont suivies de la présentation des résultats et de leur interprétation phylogénétique.

- Une troisième partie porte en effet sur la variabilité du génome chloroplastique dans son ensemble. Nous présentons sa mise en évidence à l'aide d'une approche RFLP.
- La quatrième partie porte plus particulièrement sur la variabilité de régions non-codantes du génome chloroplastique.
- La cinquième partie est consacrée à l'étude phylogénétique de l'ensemble des caractères informatifs concernant les variations de l'ADN chloroplastique.
- La sixième partie concerne le contenu en ADN nucléaire de quelques espèces. Les résultats concernant les variations interspécifiques de la taille du génome nucléaire sont discutés en relation avec la phylogénie des espèces et l'adaptation aux conditions du milieu.
- La septième et dernière partie comporte une synthèse de nos résultats sur les relations de parenté entre espèces. Nous en discutons les conséquences en rapport avec l'histoire évolutive du genre *Coffea* en Afrique et à Madagascar.

## **1. LES CAFEIERS**

## **1.1. TAXONOMIE**

Le genre *Coffea* a été créé par Linné en 1753. Il a été revu à plusieurs reprises notamment par Bentham et Hooker en 1873, de Wildeman en 1901, Lebrun en 1941, et plus récemment par Chevalier (1947), Leroy (1967) et Bridson & Verdcourt, (1988).

Chevalier, dans son étude systématique des "Caféiers du globe" (1942), citait environ soixante espèces. Depuis, ce nombre s'est enrichi grâce à la découverte de nouvelles espèces, principalement à la suite des collectes effectuées à Madagascar ainsi qu'en Afrique centrale et orientale (Bridson, 1982). Pourtant toutes les zones susceptibles de contenir des caféiers n'ont pas été prospectées. De plus certains échantillons d'herbier demeurent incomplets. La taxonomie des caféiers reste donc incomplète, même après ces nombreuses révisions.

Il n'existe aucune classification botanique de l'ensemble des caféiers. La classification la plus globale reste celle de Chevalier (1947), elle demeure donc largement utilisée bien que l'auteur lui-même ait reconnu avoir disposé de peu d'échantillons provenant d'Afrique de l'Est. Pour cette région, la systématique mise à jour par Bridson et Verdcourt est la plus largement adoptée (Bridson & Verdcourt, 1988 ; Bridson, 1994).

### **1.1.1. Les caféiers *sensu lato***

#### Une tribu....

Les caféiers appartiennent à la famille des *Rubiaceae*. Cette famille comporte plus de 10 000 espèces (dont la plupart sont ligneuses) regroupées en 630 genres (Bremer & Jansen, 1991). Les caféiers constituent la tribu des *Coffeae* (Bridson & Verdcourt, 1988). Les espèces regroupées dans cette tribu sont des arbres ou des arbustes qui portent de petites fleurs odorantes groupées par 4-5 jusqu'à 12. Les ovaires possèdent 2 loges contenant chacune un ovule. Les graines présentent un sillon ventral plus ou moins invaginé, signe distinctif de leur placentation originale, dite "cofféenne" (Leroy, 1980).

#### ...composée de 2 genres...

La tribu des *Coffeae* est subdivisée en deux genres : le genre *Coffea* L. et le genre *Psilanthus* Hook.f. (Bridson & Verdcourt, 1988). Ces deux genres se distinguent par leur structure florale particulière (Leroy, 1980; Bridson & Verdcourt, 1988). Les espèces du genre

*Coffea* ont des anthères et des stigmates externes, de plus les fleurs possèdent un style long. Au contraire les fleurs des espèces du genre *Psilanthus* possèdent des anthères et stigmates inclus dans le tube de la corolle, entourant un style court.

Chacun des deux genres est subdivisé en sous-genres, inégalement riches en espèces. Les espèces du genre *Coffea* sont classées en deux sous-genres : *Coffea* et *Baracoffea* (Leroy). Les espèces appartenant au genre *Psilanthus* sont regroupées en deux sous-genres : *Psilanthus* et *Afrocoffea* (P. Moens). La structuration botanique de la tribu des caféiers est résumée dans le tableau 1.

Genre	Sous-genre	Espèces
<i>Coffea</i>	<i>Coffea</i>	80 taxons (environ 25 espèces africaines et 55 malgaches)
	<i>Baracoffea</i>	7 espèces
<i>Psilanthus</i>	<i>Psilanthus</i>	1 seule espèce ( <i>P.mannii</i> )
	<i>Afrocoffea</i>	18 espèces

**Tableau 1.** Répartition des espèces de caféiers (tribu des *Coffeae*) d'après Bridson & Verdcourt et (1988) Bridson (1994).

### 1.1.2. Le genre *Coffea*

Le genre *Coffea* possède une aire de répartition limitée au continent africain et à Madagascar.

Les espèces du sous-genre *Coffea* (Caféiers au sens strict) sont caractérisées par des inflorescences exclusivement axillaires et une croissance monopodiale. C'est le sous-genre le plus riche en espèces. Les systématiciens en ont décrit plus de 80, dont les deux espèces de caféier cultivé : *C. arabica* et *C. canephora*. Tous les taxons appartenant au sous-genre *Coffea* sont endémiques de la zone forestière intertropicale de l'Ancien Monde (Bridson & Verdcourt, 1988; Anthony, 1992)



Les espèces constituant le sous-genre *Baracoffea* portent des fleurs terminales sur de courts "éperons" latéraux. Elles possèdent un développement de type sympodial. Seulement 7 taxons sont décrits, tous endémiques de zones côtières arides (Somalie, Kenya, Madagascar).

Chevalier avait créé un sous-genre supplémentaire (*Psilanthopsis*) (Chevalier, 1947) pour une seule espèce : *Psilanthopsis kapakata* (A. Chev.), présente exclusivement en Angola. Du point de vue botanique cette espèce présente des lobes calicinaux très développés et accrescents caractéristiques. De nombreux arguments à la fois morphologiques, cytogénétiques et biochimiques (Bridson, 1982; Louarn, 1992; Anthony *et al.*, 1993), suggéraient que l'espèce *Psilanthopsis kapakata* qui présente des affinités avec certaines espèces est-africaines du sous-genre *Coffea*, puisse être classée dans ce sous-genre. Ce reclassement a récemment été confirmé (Bridson, 1994).

### 1.1.3. Le genre *Psilanthus*

Le genre *Psilanthus* possède une aire de répartition très étendue, allant de l'Afrique à l'Océanie.

Les dix-huit espèces du sous-genre *Afrocoffea* sont réparties dans toute l'Afrique intertropicale, l'Inde, la Malaisie, et le nord de l'Australie (Bridson & Verdcourt, 1988). Ce sous-genre n'est toutefois pas représenté à Madagascar. Les fleurs de ces espèces sont le plus souvent situées en position terminale, parfois à la fois axillaire et terminale. Le mode de développement sympodial prédomine.

Le sous-genre *Psilanthus* est, lui, exclusivement présent en Afrique de l'ouest et dans le bassin du Congo (figure 1). Il est composé d'une seule espèce largement décrite (échantillon complet), *P. mannii* (Hook.f.), et de quelques autres taxons incomplets mais aisément reconnaissables à une particularité botanique des lobes du calice (Bridson & Verdcourt, 1988). Ce sous-genre est caractérisé par un mode de développement monopodial et des fleurs axillaires.

#### 1.1.4. Le genre *Coffea*, sous-genre *Coffea*

Chevalier (1942) a proposé une subdivision des *Coffea* en cinq sections : *Argocoffea*, *Paracoffea*, *Mascarocoffea*, *Mozambicoffea* et *Eucoffea*. Les espèces de la section *Argocoffea* ne possédaient pas la placentation originale des caféiers ; elles ont été éliminées du genre *Coffea* (Leroy, 1961). La section *Paracoffea* correspond aujourd'hui au genre *Psilanthus*, sous-genre *Afrocoffea* (Bridson, 1987). Par contre les trois sections restantes correspondent à l'actuel genre *Coffea* sous-genre *Coffea*. Les *Mascarocoffea* sont les caféiers de la région malgache, les *Mozambicoffea* les caféiers spontanés à l'est de la dorsale Nil-Congo et les *Eucoffea* les caféiers spontanés en Afrique de l'ouest et centrale, jusqu'à la dorsale Nil-Congo. La section des *Eucoffea* est elle-même divisée en 4 sous-sections à partir de critères morphologiques :

- La couleur des fruits à maturité ; rouge pour les *Erythrocoffea*, noire pour les *Melanocoffea*.
- La taille des arbres à l'âge adulte ; importante pour les *Pachycoffea*, réduite pour les *Nanocoffea*.

Ces subdivisions taxonomiques s'adaptent parfois mal au foisonnement de formes que possèdent les caféiers issus de prospections plus récentes. Elles sont d'ailleurs contestées, même du point de vue de la taxonomie par Bridson et Verdcourt (1988), qui eux ne proposent aucune subdivision en dessous du sous-genre.

La classification de Chevalier est basée sur des caractéristiques géographiques et de morphologie florale. Berthaud (1984), utilisant la terminologie de Chevalier (1942) et la classification de Leroy (1980), distingue sur une base géographique trois groupes de caféiers dans le sous-genre *Coffea* : *Mascarocoffea*, *Mozambicoffea* et *Erythrocoffea*.

-Les *Mascarocoffea*, comme dans la systématique de Chevalier, sont les caféiers endémiques de la région malgache (Madagascar et les îles Mascareignes).

-Les *Mozambicoffea* regroupent les caféiers endémiques de la partie est et sud-est de l'Afrique, à l'exception de deux espèces, *C.eugenioides* et *C.arabica*. Les caféiers de ce groupe présentent une durée de maturation des fruits très courte (quelques semaines à quelques mois) et un taux de caféine très bas, voire nul.

-Les *Erythrocoffea* (nom donné par Chevalier aux caféiers ayant des fruits rouges à maturité), comprennent dans le découpage de Berthaud, tous les caféiers de la zone guinéo-congolaise ainsi que les deux espèces *C.eugenioides* et *C.arabica* (situés en bordure est de cette zone) (Berthaud, 1984). Dans cette section sont regroupés donc des espèces ayant des fruits de couleur vive, caractéristique adaptative qui aurait favorisé leur transport par les

oiseaux (Berthaud, 1984). Il faut toutefois noter que l'espèce *C.stenophylla*, classée dans ce groupe possède des fruits noirs à maturité. C'est également le cas de certains taxons récemment prospectés en Afrique centrale, et donc susceptibles d'être classés dans ce groupe (Anthony, 1992).

Louarn (1992) a également proposé une structuration des espèces du sous-genre *Coffea*. Les caféiers diploïdes du sous-genre *Coffea* sont subdivisés en trois groupes selon la fertilité de leurs hybrides inter-spécifiques. Il les nomme comme Chevalier (1942) : *Mascarocoffea*, *Mozambicoffea* et *Eucoffea*. Ces sections correspondent respectivement aux caféiers de la région malgache, de l'est et du sud-est africain et à ceux originaires de l'ouest africain. Comme Berthaud, il propose que l'espèce *C.eugenioides*, en dépit de son origine kenyane, soit rattachée au groupe des caféiers de l'ouest (appelé ici *Eucoffea*). Selon lui cette espèce pourrait toutefois préfigurer une ligne de différenciation originale. Par ailleurs, l'espèce *C.kapakata* n'exprime pas de différenciation particulière et sa classification dans un sous-genre différent (sous-genre *Psilanthopsis*) ne semble pas se justifier, elle est donc classée parmi les *Eucoffea* (Louarn, 1992).

Anthony (1992) propose une organisation du complexe des caféiers en compartiments. Son découpage (indépendant des divisions taxonomiques citées précédemment) est basé à la fois sur des données concernant la répartition de la diversité et l'étude des flux de gènes. Les caféiers du sous-genre *Coffea* sont classés dans 5 compartiments :

- un compartiment pour l'Afrique occidentale,
- un compartiment pour l'Afrique centrale, scindé en 3 sous-compartiments (*C.kapakata*, *C.sp.*Moloundou, et autres *Coffea*),
- un compartiment spécifique pour l'espèce tétraploïde *C. arabica*,
- un compartiment pour l'Afrique orientale, divisé en 4 sous-compartiments sur des critères biochimiques et physiologiques,
- un compartiment pour Madagascar divisé en 2 sous-compartiments selon la teneur en caféine ( 3 taxons de la série *Verae* d'une part et autres *Coffea* d'autre part)

## 1.2. LES PRINCIPALES ESPECES

Afin de simplifier la présentation des caractéristiques des principales espèces de caféiers, nous les avons classées en quatre groupes géographiques. Les localisations des principales espèces sauvages appartenant au sous-genre *Coffea* (A,B,C) sont représentées sur la figure 1.

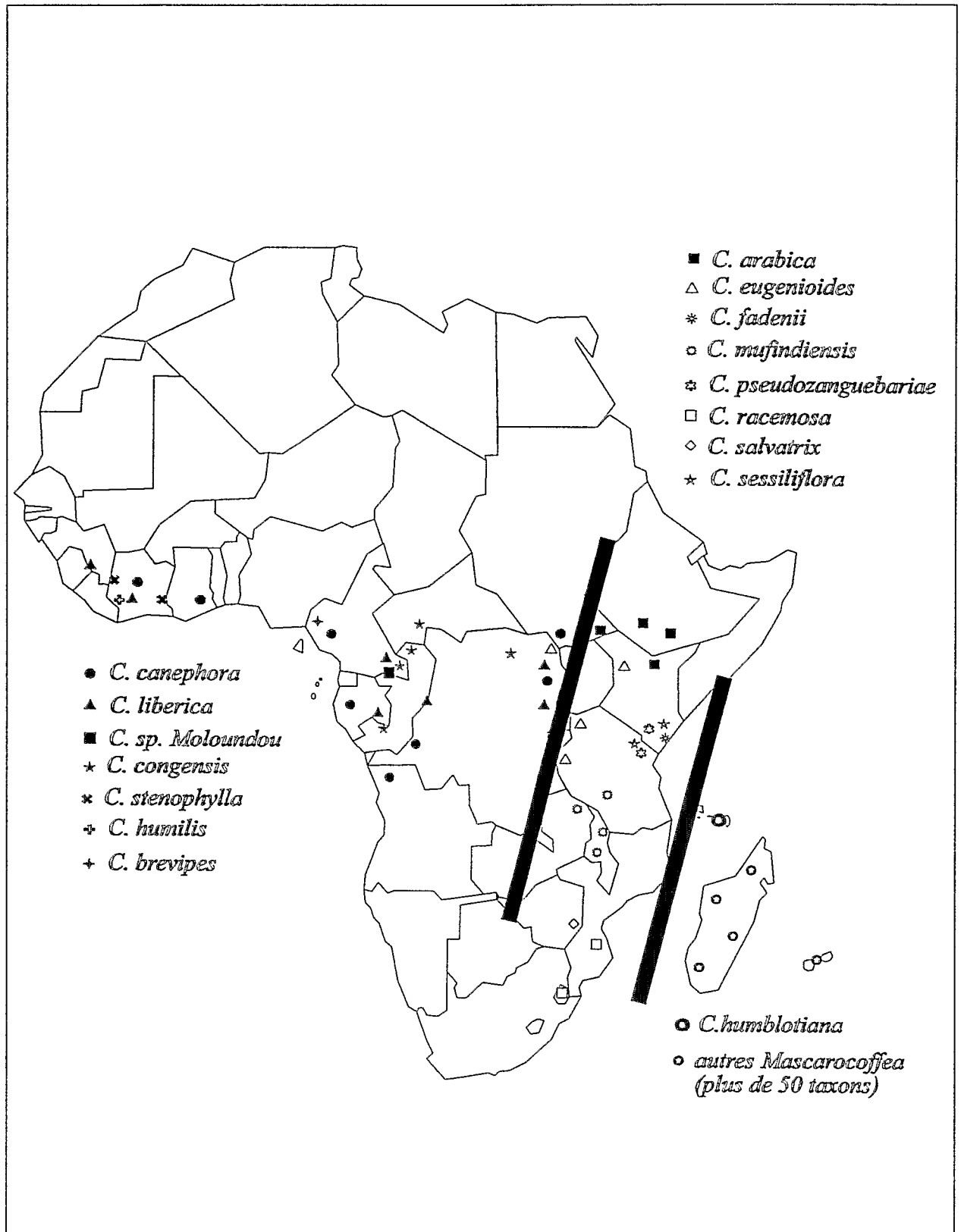
- A. Espèces du genre *Coffea* sous-genre *Coffea* originaires d'Afrique centrale et de l'ouest
- B. Espèces du genre *Coffea* sous-genre *Coffea* originaires d'Afrique de l'Est.
- C. Espèces du genre *Coffea* sous-genre *Coffea* endémiques de la région malgache.
- D. Espèces du genre *Psilanthus* sous-genre *Psilanthus* et sous-genre *Afrocoffea*.

### **1.2.1. *Coffea* sous-genre *Coffea* d'Afrique centrale et de l'ouest**

La zone intertropicale de l'Afrique de l'ouest et de l'Afrique centrale constitue la région guinéo-congolaise. Cette zone est constituée d'une large bande autour de l'équateur, à partir du littoral atlantique jusqu'au versant occidental de la dorsale du Kivu, en passant par le bassin du Zaïre (zone congolaise). Une zone satellite recouvre la région qui va de la Guinée au Ghana (zone guinéenne). Ces deux zones sont séparées par le couloir du Dahomey, beaucoup plus aride. La zone guinéo-congolaise correspond essentiellement à une forêt dense et humide (de Namur, 1990). L'altitude est inférieure à 1000 m. Dans la région du bassin du Zaïre elle est en moyenne de 400 m alors qu'elle atteint 800 m sur les plateaux du Cameroun.

Toutes les espèces classées dans ce groupe sont diploïdes avec  $2n=2x=22$  chromosomes. Elles présentent des caractéristiques morphologiques, phénologiques et biochimiques particulières. Ce groupe réunit des taxons ayant le plus souvent un port arboré, possédant des feuilles plutôt grandes et peu épaisses, de gros fruits (Anthony, 1992), et un nombre d'inflorescences par noeud généralement élevé (3 à 8) (Vishveshwara & Ahmed, 1975). Du point de vue phénologique, la caractéristique la plus remarquable concerne leur cycle de fructification qui est d'une durée de 7 à 15 mois. La teneur en caféine de leur grain varie entre 0,5 et 3,5% de la matière sèche et la teneur en acides chlorogéniques est elle aussi élevée (4,5 à 10%)

Le tableau 2 synthétise les données concernant la distribution et quelques caractéristiques botaniques et génétiques des principaux taxons constituant le groupe des espèces guinéo-congolaises.



**Figure 1.** Localisation géographique des principales espèces du genre *Coffea*, sous-genre *Coffea* (Bridson, 1982 ; Anthony, 1992). Les lignes noires figurent la limite approximative des trois zones biogéographiques.

Taxons	distribution		caractéristiques particulières	mode de reproduction	caféine/acide chlorogénique en % de M.S.
<i>C.brevipes</i> Hiern	Afrique centrale (Cameroun)			allogame	1,9 à 2,2/6,4
<i>C.canephora</i> Pierre	toute la zone guinéo-congolaise	nombreuses adaptations (0 à 1300 m) , type congolais, type guinéen, caféier de la Nana...	port multicaule,	allogame	1,6 à 3,2/6,7 à 9,2
<i>C.congensis</i> Froehner	Afrique centrale	endémique des bords de fleuves (altitude 300 à 450 m)	buisson grêle et multicaule (1 à 4 m de hauteur)	allogame	2,1/8,1
<i>C.humilis</i> A. Chev.	Afrique de l'ouest	il existe peut être une forme affine en Afrique Centrale (Berthaud, 1984)		allogame	2/6,2
<i>C.liberica</i> Hiern	toute la zone guinéo-congolaise	deux formes sont différenciées, une guinéenne ( <i>liberica</i> ) et une congolaise ( <i>dewevrei</i> )		allogame	1,8 à 2/7,3 à 7,6
<i>C.stenophylla</i> G. Don	Afrique de l'ouest		petits fruits (15mm) noir à maturité	allogame	0,6 à 2,6/5 à 9,3
<i>C.sp.Moloundou</i>	Afrique centrale	plusieurs populations au Cameroun et au Congo	petite taille, port en boule, petit fruits rouge vif à maturité	autogame	0,6/4,6

**Tableau 2.** Principales espèces de caféiers sauvages originaires d'Afrique de l'ouest (zone guinéenne) et d'Afrique centrale (zone congolaise) ; description de quelques caractéristiques morphologiques (Chevalier, 1942), du mode de reproduction (Anthony, 1992) et de leur teneur en caféine et acides chlorogéniques (Anthony *et al.*, 1993).

### 1.2.2. *Coffea* sous-genre *Coffea* d'Afrique de l'Est

Se sont des espèces endémiques d'une région délimitée à l'ouest par la dorsale du Kivu et à l'est par le canal du Mozambique (les îles côtières de Tanzanie comprises) et qui s'étend du sud de l'Ethiopie jusqu'à la province du Natal (Afrique du Sud). Cette zone possède un relief très varié : des plaines côtières du Mozambique jusqu'aux hauts plateaux du sud de l'Ethiopie ou des environs du lac Victoria. Contrairement à la zone guinéo-congolaise peuplée de forêts hautes, denses et humides, cette région phytogéographique correspond essentiellement à des forêts sèches plus ou moins basses (de Namur, 1990).

Les espèces classées dans ce groupe à l'exception de *C.arabica* et dans une moindre mesure *C.eugenoides*, présentent des caractéristiques morphologiques, phénologiques et biochimiques communes. A l'exception de *C.arabica*, du point de vue morphologique les taxons de ce groupe possèdent plutôt un port arbustif, voire buissonnant (Anthony, 1992), peu

d'inflorescences par noeud (1 à 4) (Vishveshwara & Ahmed, 1975), de petites feuilles et de petits fruits (Anthony, 1992). L'aspect buissonnant et la taille réduite des feuilles, comme la faible longueur de leurs cycles de fructification (1 mois et demi à 4 mois), semblent être en partie liés à une adaptation écologique de certaines espèces à des conditions relativement arides (longue saison sèche). De plus, par rapport aux espèces guinéo-congolaises, la teneur en caféine est très réduite (de 0 à 2% de la matière sèche) (Anthony, 1992).

Les principales espèces ayant pour origine cette zone géographique sont brièvement présentées dans le tableau 3.

Taxons	Distribution géographique (altitude)	caractéristiques particulières	mode de reproduction	caféine/acide chlorogénique en % de M.S.
<i>C.arabica</i> L.	Plateau abyssinien (1300 à 2000m)	tétraploïde (2n=44), plus de 50 variétés cultivées dérivées de 2 introductions ( <i>bourbon</i> et <i>typica</i> )	autogame	0,8 à 1,8/5,2 à 7,5
<i>C.eugenioides</i> S. Moore	Ouganda, Kenya et Tanzanie (1050 à 2200m)		allogame	0,4 à 0,6/4,5 à 6,3
<i>C.fadenii</i> Bridson	Taïta Hills, au Kenya (1400 à 1750m)	arbre pouvant atteindre 15m de hauteur	allogame	??
<i>C.mongensis</i> Bridson	District de Lushoto en Tanzanie (1100 à 1400m)		allogame	??
<i>C.mufindiensis</i> Bridson	Tanzanie, Zambie et Mozambique (1600 à 2100m)		allogame	1,3/7,2
<i>C.pseudozanguebariae</i> Bridson	zone côtière du Kenya et de la Tanzanie (jusqu'à 800m)		allogame	<0,1/0,9 à 1,8
<i>C.racemosa</i> Lour.	sud Mozambique, sud Zimbabwe, nord de l'Afrique du sud		allogame	0,8 à 1,2/4,9 à 6
<i>C.salvatrix</i> Swynnerton & Phillipson	Mozambique, Zimbabwe, et peut être Malawi (800 à 1700m)		allogame	<0,1 à 0,3/1,7 à 2,3
<i>C.sessiliflora</i> Bridson	zone côtière du Kenya et de la Tanzanie (450m max.)		allogame	0,4 à 0,7/5,6 à 9,8
<i>C.costatifructa</i> Bridson	Tanzanie et îles voisines (Mafia) (proche de 0)	arbre de taille >10m	allogame	1/9,6

**Tableau 3.** Principales espèces de caféiers originaires d'Afrique de l'est. Leur aire de répartition (Bridson & Verdcourt, 1988), quelques caractéristiques (Bridson & Verdcourt, 1988; Anthony, 1992) et leur teneur en caféine et acides chlorogéniques (Anthony *et al.*, 1993) sont précisées.

### 1.2.3. *Mascarocoffea*

Les *Mascarocoffea* sont les espèces du sous-genre *Coffea* endémique à Madagascar et les îles voisines. Leur diversité morphologique est impressionnante. Elle proviendrait d'une diversification par cladogenèse, faite d'isolements successifs (Charrier, 1976) liés au milieu physique très accidenté et aux nombreux microclimats présents sur la grande île (Anthony, 1992). Leur classification taxinomique reste assez complexe. A ce jour on compte 56 espèces regroupées en 6 séries botaniques au sein de la section des *Mascarocoffea*. Cette classification est basée sur des caractères de morphologie des feuilles, des fleurs et des fruits, complétée par la prise en compte de la phénologie et de caractères agronomiques et biochimiques (Charrier, 1976).

Les *Mascarocoffea* sont tous diploïdes et allogames et ont la particularité de ne pas posséder de caféine, ou seulement des traces (0,8% de la matière sèche maximum chez 3 espèces de la série *Verae*) (Rakotomalala, 1992; Clifford *et al.*, 1991). Bien qu'ils présentent une diversité morphologique et phénologique importante, l'obtention d'hybrides interspécifiques est possible dans la majorité des cas, ce qui traduit des divergences chromosomiques faibles (Charrier, 1976).

Nous présentons dans le tableau suivant (4) la classification en 6 séries donnée par Charrier, qui synthétise les résultats de ses propres observations et le groupement en séries définies par Chevalier (1942). Il faut toutefois noter que les caractéristiques biochimiques de chaque série ne sont pas homogènes, et qu'il est possible de les subdiviser en groupes biochimiques nettement différenciés (Rakotomalala, 1992).



Série botanique	principaux taxons	quelques caractéristiques				
		distribution géographique	feuilles	inflorescences	fruits	granulométrie
<b>Verae</b>	<i>C.lancifolia</i> Chev., <i>C.homollei</i> Leroy, <i>C.kianjavatensis</i> Leroy	domaine de l'est (plaine côtière de l'est, du nord au sud)	moyennes ou grandes, limbe souple et gauffré	uniflores, exceptionnellemen t pauciflores	petits, oblong à ovoïdes	3 à 6 g /100 graines
<b>Mauritianae- Humblotiana</b>	<i>C.humblotiana</i> Baillon, <i>C.mauritiana</i> Lamk., <i>C.macrocarpa</i> Rich. , <i>C.vaughanii</i> Leroy	Comores et îles Mascareignes	petites ou moyennes	uniflores, parfois pauciflores	moyens, ovoïdes parfois très allongés	10 à 18 g /100 graines
<b>Multiflorae (élargie)</b>	<i>C.perrieri</i> Drake, <i>C.resinosa</i> Hook. f., <i>C.vianneyi</i> Leroy, <i>C.bertrandi</i> Chev. ...	Partout à Madagascar, mais essentiellement sur la moitié est	grandes (10 à 15 cm), coriaces	pluriflores plus ou moins retombantes	ovoïdes allongés	3 à 8 g /100 graines
<b>complexe Milotii</b>	<i>C.milotii</i> Leroy, <i>C.dolichophylla</i> Leroy, <i>C.farafanganensis</i> Leroy, <i>C.richardii</i> Leroy...	domaine de l'est (plaine côtière de l'est, du nord au sud)	grandes (10 à 30 cm), épaisses et coriaces	pluriflores	forme et taille de prunes	18 à 30 g /100 graines
<b>Garcinioides</b>	<i>C.dubardi</i> Jumelle, <i>C.mogeneti</i> Dubard, <i>C.tetragona</i> Jumelle & Perrier, <i>C.heimii</i> Leroy...	région du nord et côte nord ouest	moyennes (5 à 11 cm), formes variées	pluriflores, dressées	taille moyenne oblongs à arrondis	5 à 17 g 100 graines
<b>Subterminales</b>	<i>C.pervilleana</i> Drake, <i>C.sakarahae</i> Leroy, <i>C.vatovavyensis</i> Leroy, <i>C.tsirananae</i> Leroy...	région du nord en majorité, quelques taxons de côte est	petites (2 à 6 cm)	uni ou pluriflores, axillaires sur les noeuds terminaux ou subterminaux	ovoïdes à oblongs	3 à 8 g 100 graines

**Tableau 4.** Synthèse des principaux résultats concernant la classification des *Mascarocoffea* en 6 séries, d'après (Charrier, 1976).

#### 1.2.4. Le genre *Psilanthus*

Le genre *Psilanthus* est voisin du genre *Coffea* du point de vue taxonomique. Contrairement au genre *Coffea*, il possède une aire de répartition très large, allant de l'ouest de l'Afrique à l'Océanie en passant par l'Inde, mais il n'est pas présent à Madagascar. Les taxons de ce genre semblent tous autogames ce qui rend improbable l'existence de flux de gènes entre les représentants des 2 genres. Toutefois en condition artificielle il a été récemment possible d'obtenir des hybrides inter-génériques entre *Coffea* et *Psilanthus* (Couturon, communication personnelle).

Ce genre est divisé en deux sous-genres sur la base du mode de développement. Quelques espèces fréquentes sont citées dans le tableau ci-dessous :

Sous-genre	Espèce	distribution géographique	caractéristiques	caféine/acides chlorogéniques en % de M.S.
<b>Psilanthus</b>	<i>P.mannii</i> Hook.f.	zone guinéo-congolaise		0/1,1
<b>Afrocoffea</b>	<i>P.bengalensis</i> Leroy	Inde et sud est asiatique		0,15/1
	<i>P.ebracteolatus</i> Hiern	zone guinéo-congolaise		<0,1/1
	<i>P.leroyi</i> Bridson	Ethiopie, berges de rivière (1200m)	2 à 6m de hauteur, petits fruits	??
	<i>P.semseii</i> Bridson	Tanzanie (500m)	buisson jusqu'à 2,5m de hauteur, petits fruits	??

**Tableau 5.** Quelques espèces de *Psilanthus* ; leur distribution géographique (Chevalier, 1942; Bridson & Verdcourt, 1988; Anthony, 1992) et quelques caractéristiques (Bridson & Verdcourt, 1988; Anthony *et al.*, 1989) sont précisées.

### 1.3. LES COLLECTIONS DE RESSOURCES GENETIQUES

#### 1.3.1. Prospections

La première "collecte" de caféiers spontanés sur le plateau éthiopien par certaines populations de la péninsule arabique est probablement antérieure au XV<sup>ème</sup> siècle. Elle a permis d'introduire le *C.arabica* au Yémen, qui est considéré aujourd'hui comme un centre secondaire d'origine. L'exploration et la collecte des formes sauvages de caféiers par les botanistes, qui ont ramené quelques individus dans les jardins botaniques occidentaux, n'a réellement débuté qu'à la fin du XIX<sup>ème</sup> siècle (Chevalier, 1929).

L'exploration systématique pour la collecte des formes sauvages ne débute, quant à elle, qu'au milieu du XX<sup>ème</sup> siècle. Elle a été rendue urgente par la mise en évidence de l'érosion génétique. Cette érosion est liée à la destruction rapide d'habitats forestiers tropicaux, à la dynamique du système agricole et à la surexploitation des terres. La F.A.O.<sup>1</sup> puis l'I.B.P.G.R.<sup>2</sup>

<sup>1</sup>Food and Agriculture Organisation

<sup>2</sup>International Board for Plant Genetic Ressource

respectivement dans les années 60 puis 80 inscrivent le caféier comme plante prioritaire à collecter. Ces collectes ont été menées en collaboration par des institutions internationales et des organismes de recherche caféière.

On peut considérer que 20.000 génotypes spontanés, représentant 70 taxons du sous-genre *Coffea*, ont été prospectés. La conservation de ces ressources génétiques représente un investissement important. Contrairement à de nombreuses espèces cultivées conservables sous forme de graines, les ressources génétiques caféières ne peuvent pas être conservées en banques de gènes classiques. Leurs graines (comme celles du cacao, de l'avocat ou du caoutchouc, etc...) sont dites "récalcitrantes". Leur conservation (avec maintien du pouvoir germinatif) n'est assurée que pendant 8 à 10 mois dans des conditions de température et d'hygrométrie contrôlées (25°C saturé en humidité) (An., 1991). La constitution de collections vivantes, en champ, en serre ou *in vitro* est donc indispensable.

### 1.3.2. Inventaire des collectes

L'espèce *Coffea arabica* a fait l'objet de plusieurs collectes en Ethiopie (Guillaumet & Hallé, 1978). La première initiée en 1964 par la F.A.O et plusieurs pays producteurs de café, avait pour objectif la collecte de cultivars locaux et semi-spontanés. La seconde réalisée en 1966 par l'ORSTOM et l'IRCC<sup>3</sup> s'est principalement intéressée aux formes sauvages. Depuis, un certain nombre d'autres prospections ont eu lieu ; un double du matériel collecté a été rassemblé sur la station éthiopienne de Jimma.

De 1960 à 1974 de nombreuses missions de prospection organisées par le Muséum National d'Histoire Naturelle, l'IRCC et l'ORSTOM ont eu lieu à Madagascar. La diversité des peuplements naturels est très importante. On estime que les 10.000 individus collectés appartiennent à une cinquantaine de taxons. Le matériel collecté est rassemblé sur les stations malgaches de Kianjavato et d'Ibaka-Est.

Les *Eucoffea* diploïdes ont été prospectés par l'ORSTOM et l'IRCC entre 1975 et 1987 dans sept pays africains (Côte d'Ivoire, Guinée, Congo, Cameroun, Centrafrique, Tanzanie et Kenya). Sur 230 sites prospectés au total, 8500 génotypes appartenant à une vingtaine de

---

<sup>3</sup>Institut de Recherche sur le Café et le Cacao

taxons ont pu être récoltés. Le matériel collecté est conservé d'une part dans le pays où a eu lieu la collecte et d'autre part au CRG<sup>4</sup> des caféiers africains en Côte d'Ivoire.

### 1.3.3. La collection du CRG africain

L'ORSTOM, en tant que gestionnaire du CRG des caféiers africains en Côte d'Ivoire est un des principaux organismes intervenant dans la sauvegarde des ressources génétiques caféières (Anthony, 1992). De plus, sur le Centre de Montpellier, une "core collection" est en cours d'établissement : elle doit réunir, en un nombre réduit d'introductions, la diversité maximum observable chez les caféiers. Cette "core collection" est conservée à la fois *in vitro* et en serres.

## 1.4. STRUCTURATION ET ORGANISATION DU COMPLEXE D'ESPECE

Les espèces de caféiers présentes en collection ont été étudiées pour un certain nombre de caractères morphologiques et phénologiques. Certaines espèces ont bénéficié d'études plus approfondies, isoenzymatiques et biochimiques. De plus, certains génotypes ont été utilisés dans des programmes d'hybridations interspécifiques. Toutes les données accumulées donnent une image de la diversité actuelle existante au sein de ce complexe d'espèces et permettent de proposer des hypothèses évolutives responsables de cette structuration.

### 1.4.1. Ploïdie, mode de reproduction

C'est sur le sous-genre *Coffea* qu'ont été accumulées le plus d'informations génétiques. A l'exception de *C.arabica* qui est tétraploïde, tous les autres caféiers pour lesquels des comptages chromosomiques ont été réalisés se sont avérés diploïdes avec  $2n=2x=22$  chromosomes. *C.arabica* possède de plus un mode de reproduction particulier chez les caféiers : l'autogamie préférentielle. Seuls deux autres taxons diploïdes possèdent cette caractéristique (*C.brevipes* var. *heterocalyx* et *C.sp.* Moloundou) ; les autres espèces diploïdes sont strictement allogames.

---

<sup>4</sup>Centre des Ressources Génétiques

#### 1.4.2. Diversité morphologique

Les différentes espèces sauvages de *Coffea* de la forêt tropicale humide sont de taille, de port et d'aspect très variable. La morphologie des feuilles et des fruits (couleur, taille et forme) est elle aussi assez variable. Certaines tendances de structuration liées à l'écologie de la zone d'endémisme semblent prédominer.

Les espèces de caféiers d'Afrique centrale et occidentale ont le plus souvent un port arboré, de grandes feuilles peu épaisses et de gros fruits en général de couleur claire. Au contraire, la plus grande partie des espèces de caféiers originaires d'Afrique orientale, où la période sèche est plus longue et plus marquée, sont des arbustes buissonnants à feuilles petites et épaisses, possédant de petits fruits de couleur sombre et présentant un cycle de fructification plus court que les espèces de la forêt humide (Anthony, 1992).

#### 1.4.3. Données izoenzymatiques

L'utilisation de marqueurs isoenzymatiques a permis de caractériser plus précisément ces ressources génétiques. Un certain nombre d'enzymes a été étudié, et huit systèmes sont couramment analysés :

- trois oxydoréductases (la malate déshydrogénase ou MDH, l'isocitrate déshydrogénase ou ICD et la 6-phosphogluconate déshydrogénase ou PGD),
- une transférase (la phosphoglucomutase ou PGM),
- deux hydrolases (l'estérase ou EST et la phosphatase acide ou PAC),
- une peptidase (la leucine aminopeptidase),
- une isomérase (la phosphoglucose isomérase ou PGI).

Les études réalisées n'ont chaque fois concerné qu'un nombre restreint d'espèces proches les une des autres. Plusieurs travaux ont été réalisés sur des espèces originaires d'Afrique centrale et de l'ouest, *C.canephora*, *C.congensis*, *C.liberica*, *C.eugenioides*, *C.humilis* et *C.stenophylla* (Berthou *et al.*, 1980; Berthaud, 1984; Berthaud *et al.*, 1984) ; et également sur les espèces est-africaines *C.pseudozanguebariae*, *C.sessiliflora*, *C.sp F* (Anthony, 1992). Des études plus ponctuelles et non publiées ont concerné des populations récemment prospectées en Afrique centrale (Anthony, communication personnelle).

Tous ces travaux ont donné des résultats intéressants quant à la structuration des populations au niveau intraspécifique. La présence d'allèles particuliers des enzymes PGD 2 et PGI 1 ont

permis de mettre en évidence une structuration en deux groupes, dit guinéen et congolais chez l'espèce *C.canephora* (Berthaud, 1984). De même les échantillons de *C.liberica* sont caractérisés par une distance génétique "intra pays" faible et "inter pays" forte, ce qui suggère une différenciation en deux groupes équivalents à ceux observés chez *C.canephora* (Berthou *et al.*, 1980). Berthou *et al.* (1980) retrouvent également deux populations isolées et génétiquement appauvries. Les caféiers de la Nana, hétérogènes, et morphologiquement aussi proches de *C.canephora* que de *C.congensis*, se sont, dans tous les cas, révélés très proches des *C.canephora* du point de vue isoenzymatique (Berthou *et al.*, 1980). Les caféiers *C.sessiliflora* originaires de Tanzanie (population Kitulangalo) sont nettement séparés des *C.sessiliflora* originaires du Kenya (population Shimba), en particulier par leur richesse en électromorphe dans le système MDH (Anthony, 1992).

En ce qui concerne l'analyse des relations phylogénétiques, la comparaison des distances génétiques interspécifiques est rarement opérante en raison de la valeur toujours élevée des distances génétiques intraspécifiques.

#### 1.4.4. Diversité biochimique

De nombreux composés chimiques présents dans les grains des caféiers ont été dosés : lipides, hydrates de carbone, protéines, acides aliphatiques et chlorogéniques, minéraux, alkaloïdes. Néanmoins la plupart des travaux relatifs à leur analyse concernent exclusivement les deux espèces cultivées : *C.arabica* et *C.canephora (robusta)* (Maier, 1993; Speer *et al.*, 1993; Silwar & Lüllmann, 1993). On peut aisément distinguer ces deux espèces sur la base de leur composition chimique.

En ce qui concerne la diversité observée sur les espèces de caféiers sauvages, plusieurs études ont été menées, certaines portant seulement sur quelques espèces, d'autres sur un échantillonnage beaucoup plus large (Charrier & Berthaud, 1975; Santa Ram *et al.*, 1982; Anthony *et al.*, 1989; Rakotomalala, 1992; Anthony *et al.*, 1993; Rakotomalala *et al.*, 1993; Rakotomalala *et al.*, 1993). Les composés les plus fréquemment dosés sont ceux associés au métabolisme de la caféine, qui ont été analysés, soit qualitativement par des tests biochimiques, soit quantitativement par chromatographie liquide à haute pression (HPLC). Les principaux composés répertoriés sont les acides chlorogéniques et la caféine (tout deux appartenant à la classe des bases puriques) et dont la teneur est variable à l'intérieur même d'une espèce ; et les hétérosides diterpéniques, apparemment spécifiques aux caféiers sauvages.

Les analyses mettent en évidence une composition biochimique très variable des grains de café analysés. Anthony *et al.* (1993) proposent un premier classement des caféiers en trois groupes sur la base du contenu en caféine, acides féruloylquinique et dicafféoylquinique d'une part et en mozambiosides d'autre part. Les principaux résultats sont présentés dans le tableau 6 :

	Teneur en caféine (en % de matière sèche)	Teneur en mozambiosides (en % de matière sèche)
A (15 taxons <i>Coffea</i> africains)	0,46 - 3,19	0 - 0,07
B (2 taxons <i>Coffea</i> africains)	0,02 - 0,28	2,28 - 2,71
C (8 taxons <i>Mascarocoffea</i> )	0 - 0,09	0 - 0

**Tableau 6 .** Caractérisation biochimique de trois groupes de caféiers du sous-genre *Coffea*, pour 2 composés : caféine et mozambiosides (Anthony *et al.*, 1993)  
 Groupe A : *C.arabica* , *C.brevipes* , *C.canephora* , *C.eugenioides* , *C.humilis* , *C.kapakata* , *C.liberica* , *C.mufindiensis* , *C.racemosa* , *C.sessiliflora* , *C.sp F* , *C.sp.* Moloundou , *C.sp.* Nkoumbala et *C.stenophylla*.  
 Groupe B : *C.pseudozanguebariae* et *C.salvatrix*  
 Groupe C : *C.farafanganensis* , *C.homollei* , *C.perrieri* , *C.rhamnifolia* , *C.tetragona* , *C.vatovavyensis* , *C.vianneyi* et *C.sp. A315*

Une analyse multivariée réalisée exclusivement avec les taxons du groupe A montre une structuration assez forte de ce groupe en fonction des teneurs en caféine, en acide dicafféoylquinique et féruloylquinique. Celles ci sont sensiblement plus élevées chez la plupart des espèces d'Afrique centrale et de l'ouest que chez les espèces endémiques d'Afrique de l'Est (Anthony *et al.*, 1993). La teneur en caféine peut atteindre 4% chez les espèces guinéo-congolaises alors qu'elle ne dépasse pas 2% chez celles originaires d'Afrique de l'Est (Rakotomalala, 1992).

Les espèces du genre voisin *Psilanthus* possèdent des teneurs en caféine inférieures à 0,28% de la matière sèche, ce qui les rapprocherait plutôt des groupes B ou C. Les différentes espèces (*P.mannii*, *P.ebracteolatus*, *P.bengalensis* et *P.brassii*) présentent par contre des teneurs en mozambiosides très variables (Anthony *et al.*, 1989). L'affinité biochimique de deux espèces d'Afrique de l'Est (*C.pseudozanguebariae* et *C.salvatrix*), des espèces malgaches et de certaines espèces du genre *Psilanthus* serait liée à une chaîne métabolique particulière

conduisant à des teneurs faibles en caféine (maximum 0,28% de matière sèche) et en acides chlorogéniques (2,41% au maximum) (Anthony, 1992).

La caféine est un marqueur chimiotaxonomique qui permet de caractériser deux groupes fondamentaux, caféiers contenant de la caféine (groupe A) et caféiers sans caféine (groupe B+C). La structuration de la diversité biochimique globale correspond toutefois à un continuum répondant au concept de complexe d'espèces, difficilement analysable en phylogénie. Ces marqueurs ont donc été presque exclusivement étudiés à l'aide d'analyses multivariées, pour obtenir une image de la diversité.

### 1.4.5. Hybridations interspécifiques et fertilité hybride

Les résultats d'hybridations interspécifiques et la fertilité des hybrides obtenus ont mis en évidence une diversité de comportement cytogénétique. Les observations de plaques métaphasiques montrent que les taxons du sous-genre *Coffea* dérivent tous d'un même génome de base.

Les résultats d'hybridations interspécifiques suggèrent qu'il est plus facile d'hybrider des espèces de même groupe géographique (Charrier, 1976; Louarn, 1992) alors que l'obtention d'hybrides entre des espèces appartenant à des groupes biogéographiques différents reste difficile (Louarn, 1992). Au sein des *Mascarocoffea* par exemple, les hybridations interspécifiques produisent des plantes très fertiles, aucune barrière reproductive ne semble exister (Charrier, 1976). Les résultats obtenus par croisements interspécifiques contrôlés ne sont donc pas toujours en accord avec les limites entre espèces définies par les taxinomistes. Ainsi, le taux de réussite des hybridations interspécifiques est très variable en fonction des espèces concernées. Il peut quelquefois atteindre des valeurs voisines de celles obtenues en hybridation intraspécifique, par exemple dans le cas de croisements *C.canephora* (femelle) X *C.congensis* (mâle) qui donnent jusqu'à 85 hybrides pour 100 fleurs fécondées ; ou au contraire être très faible comme dans le cas de croisements *C.canephora* (femelle) X *C.stenophylla* (mâle), deux espèces sympatriques chez qui on observe 2 hybrides seulement pour 27 fleurs fécondées (Louarn, 1993).

L'ensemble des résultats obtenus par Louarn (1993) concernant la viabilité pollinique des hybrides F1 lui permet de proposer la structuration des *Eucoffea* diploïdes en 3 groupes :

-le groupe "ouest" constitué de *C.canephora*, *C.congensis*, *C.liberica* (*liberica* et *dewevrei*), *C.humilis* et *C.stenophylla*;



-le groupe "est et sud-est" comprenant *C.pseudozanguebariae*, *C.racemosa*, *C.sessiliflora* et *C.salvatrix*;

-le "groupe"centre est" constitué par la seule espèce d'altitude *C.eugenioides*.

Ce type d'expériences permet d'expliquer partiellement les relations cytogénétiques entre espèces et de mettre en évidence les barrières reproductives. Il faut toutefois analyser les résultats d'étude des relations méiose-fertilité avec précaution car la mise en évidence de différences peut être liée à des caractères ne reflétant pas la différenciation génétique globale entre deux entités. L'obtention aisée d'hybrides interspécifiques est indéniablement le signe d'une proximité des génomes. Par contre le fait de ne pas obtenir d'hybrides interspécifiques peut être lié à la seule présence d'un gène de stérilité.

## **1.5. L'HISTOIRE DES CAFEIERS AU SEIN DES FLORES AFRICAINE ET MALGACHE**

La diversité observable aujourd'hui est directement reliée (et fiable) à l'histoire et à l'évolution du continent africain. Pour exploiter les données concernant la diversité des caféiers et proposer des hypothèses sur les voies de spéciation, nous nous baserons sur les informations concernant la formation géographique de l'Afrique et l'évolution de sa couverture végétale.

### **1.5.1. L'évolution du milieu physique**

L'histoire de l'Afrique est étroitement liée à celle du Gondwana, issu de la première fragmentation de la Pangée vers 1750 Ma. A la fin de l'ère primaire, le Gondwana forme encore un bloc continental d'un seul tenant. Vers 300 Ma il subit une très importante glaciation qui inaugure une histoire commune à l'Afrique, l'Amérique du sud, l'Australie l'Inde et Madagascar. La dislocation de ce continent commence par l'ouverture de l'océan Atlantique vers 100 Ma, et se poursuit avec le départ du bloc Inde/Madagascar, puis la séparation de l'Inde. La scission entre l'Afrique continentale et Madagascar aurait eu lieu vers 80 Ma. L'Inde péninsulaire, coupée de Madagascar, subit un déplacement vers le nord, qui se terminera par une série de collisions avec l'Asie (formation de l'Himalaya). La dislocation de l'ancien Gondwana se prolonge au miocène et jusqu'à aujourd'hui, par la séparation de l'Afrique et de l'Arabie par la mer Rouge (Raven & Axelrod, 1974).

La migration "directe" entre Madagascar et l'Afrique continentale a été possible jusqu'à environ 80 Ma, mais de nombreuses espèces végétales et animales ont apparemment traversé le canal du Mozambique bien après le milieu du crétacé (Raven & Axelrod, 1974). Au trias et au quaternaire, Madagascar semble avoir été partiellement recouvert sur ses bordures ouest et sud par des transgressions dont les sédiments sont les témoins. Le canal du Mozambique a donc été moins large qu'il ne l'est aujourd'hui. De plus, de nombreux cycles d'érosion, des fractures, du volcanisme ont en permanence modifié les reliefs. L'activité volcanique intense et récente (entre 2 et 8 Ma) dans la région malgache est également à l'origine d'îles : l'archipel des Comores et celui des Mascareignes.

Les événements les mieux connus sont ceux survenus pendant l'ère quaternaire (entre 0 et 2,5 Ma). On estime par exemple, qu'au cours de la dernière grande glaciation du quaternaire (vers 18000 ans BP), le volume des calottes glaciaires était de 80 millions de km<sup>3</sup>. Ceci impliquerait une baisse du niveau de la mer d'au moins 120 m correspondant à un rétrécissement du canal du Mozambique de plusieurs dizaines de kilomètres. La végétation des îles volcaniques récentes, ainsi que le rétrécissement périodique de la largeur du canal du Mozambique, laissent supposer que dans de nombreux cas la migration pendant le quaternaire a été possible entre le continent africain, Madagascar et les îles Mascareignes.

La succession d'épisodes glaciaires au cours du quaternaire s'est traduite dans la zone intertropicale par des alternances de périodes pluviales et de périodes modérément sèches, accompagnées par de fortes variations de température. On estime qu'il y a eu entre 10 et 20 épisodes glaciaires. Chacun a duré au moins 100.000 ans, réparti en un refroidissement assez lent (étalé sur 50.000 ans) et une déglaciation rapide (moins de 10.000 ans). Sur le Kilimanjaro, les moraines sont descendues jusqu'à plus de 2.000 m en dessous de leur niveau actuel. Ces variations climatiques ont eu une influence importante sur la couverture végétale. Au cours des phases arides les taxons forestiers ont pu se maintenir dans des zones privilégiées (refuges). Ils se dispersaient à nouveau lorsque les conditions redevenaient plus humides et plus chaudes. En altitude, cette forte pluviométrie était alors responsable d'une érosion importante surtout lorsque la couverture végétale était encore éparse.

### 1.5.2. L'évolution de la flore

Selon des données moléculaires récentes, établies à partir de séquences de gènes nucléaires et chloroplastiques, la séparation entre les Gymnospermes et les Angiospermes daterait de la même époque que la séparation entre la Laurasia et la Gondwanie (330 +/- 40 Ma) (Martin *et al.*, 1993). Une origine ouest gondwanienne des Angiospermes a été suggérée. Elle aurait été suivie d'une radiation initiale lorsque la migration directe était possible entre Amérique du sud, Afrique, Inde, Antarctique, Australie et Laurasia via l'Afrique (Raven & Axelrod, 1974). La séparation entre les Mono- et les Dicotylédones daterait, elle, de 300 +/- 20 Ma (Martin *et al.*, 1993). Cette hypothèse est basée sur le concept, largement controversé, d'horloge moléculaire et les estimations divergent de façon importante. En effet une estimation basée sur la séquence codante de 12 gènes chloroplastiques et de 2 gènes ribosomiques nucléaires, place cette divergence à 200 Ma; une autre basée exclusivement sur la séquence de gènes nucléaires, place cette divergence à 320 Ma (Clegg, 1990). L'horloge moléculaire absolue n'existe pas, il est donc parfois difficile de dater avec précision les embranchements évolutifs.

De nombreuses espèces de Bryophytes sont retrouvées à la fois à Madagascar et en Afrique de l'est (Pocs, 1975). L'analyse de leur distribution montre que ce sont les espèces présentes sur les vieilles chaînes du précambrien du Kenya et de Tanzanie qui sont les plus souvent présentes à Madagascar (17,8 % d'espèces de ces origines sont communes aux deux zones). Selon Pocs, la séparation de l'Afrique et de Madagascar aurait eu lieu aux environs de 100 Ma, "juste après" la séparation entre Cryptogames et Phanérogames. Ceci expliquerait selon lui que l'on retrouve plus de Bryophytes et autres Cryptogames que de Phanérogames en commun entre Madagascar et l'Afrique continentale.

### 1.5.3. Le cas des caféiers

Toutes les propositions actuelles sur l'origine et les voies de spéciation des caféiers sont basées sur l'observation de la diversité (à l'aide de différents marqueurs) des espèces contemporaines. Nous avons montré au paragraphe précédent que la diversité génétique des caféiers semble assez bien correspondre à la répartition dans trois régions phytogéographiques : guinéo-congolaise, est-africaine, et enfin Malgache (Madagascar et îles de l'océan indien).

La présence d'espèces apparentées de part et d'autre du canal du Mozambique (*C. grevei* et *C. paolia*, 2 espèces du sous-genre *Baracoffea*), a permis d'émettre l'hypothèse d'une origine

kenyane des caféiers précédant une radiation à Madagascar (Leroy, 1982). Les *Mascarocoffea* auraient été isolés à Madagascar en une vague primitive. Selon Leroy cette hypothèse implique que l'invasion de Madagascar par les espèces ancêtres de *Coffea* soit antérieure à la dislocation gondwanienne (80 à 100Ma). Toutefois la découverte récente de taxons biochimiquement affines des *Mascarocoffea* en Afrique de l'Est repose la question de la colonisation de Madagascar (Anthony, 1992; Rakotomalala, 1992). S'agit-il réellement d'une vague primitive aux environs de 100 Ma, ou plutôt d'une colonisation plus récente rendue possible par un rétrécissement du canal du Mozambique au cours du quaternaire ? Selon Anthony (1992), la colonisation ancienne du continent vers Madagascar, aurait été complétée par des migrations plus récentes dans les deux sens, qui expliqueraient une très forte ressemblance biochimique entre certains taxons malgaches et tanzaniens (Rakotomalala, 1992).

Le genre *Coffea* présente une diversité morphologique importante à Madagascar. D'autres genres non endémiques y sont également représentés par un nombre d'espèces beaucoup plus élevé que partout ailleurs (Guillaumet & Mangenot, 1975). Ceci semble dû à la diversification par cladogénèse qui induit à Madagascar un foisonnement de formes plus ou moins nettement spécifiées (Anthony, 1992).

La différenciation des espèces africaines semble particulièrement liée à l'histoire des glaciations du quaternaire et des refuges forestiers (Berthaud, 1984; Anthony, 1992). Berthaud propose une superposition de deux cartes : l'une représentant les refuges forestiers probables au cours de la dernière grande glaciation du quaternaire (-18000 ans), l'autre la répartition actuelle des espèces de caféiers (Berthaud, 1984). La région guinéo-congolaise présente trois centres comme cela avait par ailleurs été suggéré (White, 1979). Dans la zone tanzano-mozambicaine la diversité semble plutôt liée à une mosaïque de petits îlots forestiers. Selon Anthony (1992) les taxons endémiques d'Afrique orientale se sont réfugiés dans deux zones principales, au nord au Kenya et en Tanzanie (une zone de plaine et plusieurs refuges d'altitude) et au sud vers la frontière Zimbabwe-Mozambique. L'abondance de microclimats et la position charnière entre Madagascar et la zone guinéo-congolaise seraient à l'origine de la richesse en espèces de la zone tanzano-mozambicaine.

Sur la base d'arguments génétiques (croisements contrôlés, travaux de cytogénétique...) Berthaud propose une première séparation entre les groupes de caféiers à l'intérieur du Kenya (et non pas entre l'Afrique et Madagascar, ou entre les 2 bords du Rift). Les taxons endémiques de la zone guinéo-congolaise proviendraient d'une première vague de migration des caféiers est-africains par l'axe méridional, à travers le Zaïre, l'Angola et le Congo. Ils auraient occupé toute la zone guinéo-congolaise et trouvé refuge dans trois zones au cours des

périodes arides du pléistocène : Haute-Guinée, Cameroun-Gabon et Zaïre oriental. Les refuges de Haute Guinée et du Cameroun-Gabon sont séparés par le fleuve Niger et la Cross River, qui marquent la discontinuité entre zone guinéenne et zone congolaise (Anthony, 1992). L'histoire de certains taxons endémiques de cette zone (*C.kapakata* et *C.sp.Moloundou*) pourrait néanmoins être plutôt liée à celle des taxons montagnards orientaux.

## **1.6. BILAN**

Les caféiers appartiennent à la famille des Rubiacées, au sein de laquelle ils forment la tribu des *Coffeae*, composée de 2 genres : *Coffea* et *Psilanthus*. Chacun de ces genres est subdivisé en sous-genres inégalement riches en espèces. Les espèces actuellement cultivées appartiennent toutes au genre *Coffea* sous-genre *Coffea*.

Les généticiens décrivent trois groupes ou "ensembles biogéographiques" à l'intérieur du sous-genre *Coffea*. Ce découpage est soutenu par des résultats concernant la diversité morphologique, biochimique et cytogénétique. La position de certaines espèces (*C.kapakata*, *C.eugenioides*, *C.arabica*) demeure incertaine.

Le sous-genre *Coffea* est constitué d'environ 80 taxons, endémiques de la zone intertropicale africaine et malgache. Le foisonnement taxonomique reflète une forte variabilité des caractères morphologiques. Quelle est la signification réelle du découpage en espèces ? Peut-on proposer des relations phylogénétiques entre elles ? Peut on proposer des hypothèses à l'origine de la seule espèce tétraploïde *C.arabica* ?

Les nombreuses données accumulées placent les caféiers en situation privilégiée pour établir des modèles concernant l'évolution des espèces pérennes dans la zone africano-malgache.

## **2. L'ADN CHLOROPLASTIQUE POUR LA RECONSTRUCTION DES PHYLOGENIES**

La reconstruction phylogénétique passe par la compréhension des processus évolutifs qui génèrent la diversité biologique. Au cours des dernières décennies de nombreux caractères ont été utilisés pour l'étude des relations phylogénétiques : des caractères morphologiques, cytogénétiques, puis les protéines, les composés biochimiques et enfin les variations de l'ADN (Darlu & Tassy, 1993). Cette dernière famille de caractères permet de mettre en évidence (directement ou indirectement) des différences de séquence nucléotidique. Les avantages intrinsèques de ces caractères moléculaires, et les méthodes d'analyse envisageables seront mis en évidence dans le premier paragraphe de ce chapitre.

Quel que soit le type de caractère choisi (morphologique, biochimique, moléculaire...), l'étape initiale de toute étude visant à établir des relations phylogénétiques consiste à trouver le marqueur adéquat : c'est à dire un marqueur qui varie entre groupes à étudier (populations, espèces, genres, familles) mais pas à l'intérieur d'un groupe, de façon à établir des liens de parenté entre ces groupes. Pour que la phylogénie établie à partir de ce marqueur ait un sens du point de vue de la phylogénie des espèces considérées, le marqueur choisi doit posséder un taux d'évolution compatible avec le taux d'évolution des espèces. Ce marqueur doit de plus fournir des données aisément analysables.

Chez les végétaux, c'est l'ADN chloroplastique (ADNcp) qui est le plus utilisé pour étudier les relations phylogénétiques. Après une brève revue concernant la structure et les mécanismes de l'évolution de l'ADNcp, nous aborderons ses potentialités et les différentes approches envisagées pour la reconstruction phylogénétique principalement au niveau intragénérique.

## **2.1. PHYLOGENIES MOLECULAIRES**

### **2.1.1. Caractéristiques communes aux caractères moléculaires**

Les caractères moléculaires sont de très bons marqueurs génétiques : c'est à dire qu'ils révèlent directement la nature génétique de l'information. Ceci les distingue des caractères morphologiques, physiologiques et plus généralement de tous les caractères phénotypiques, dus à une addition d'effets génétiques et non-génétiques. Cette qualité génétique leur confère un large avantage du point de vue de la reconstruction de phylogénies, par rapport aux caractères classiques utilisés en systématique. Cet avantage est traduit par deux facteurs :

- Ils sont héritables,
- ils sont totalement indépendants des effets externes (stade de développement, environnement...).

Contrairement à des caractères quantitatifs, les caractères moléculaires sont des caractères à plusieurs états finis (2 ou 4 maximum). On observe deux états possibles dans le cas de sites de restriction, codés en binaire 0/1 pour l'absence ou la présence du site. Dans le cas d'analyse directe de séquence, ce sont au maximum quatre états correspondant aux quatre nucléotides A,C,G,T, qui peuvent être codés. Les caractères moléculaires sont donc naturellement discrets (ce qui n'est pas le cas des caractères biochimiques par exemple). Du point de vue de la phylogénie, cela signifie que l'on peut employer pour reconstruire l'arbre, soit une analyse de caractères (comme pour les caractères morphologiques discrets) soit une méthode quantitative (comme dans le cas de données continues). Les caractères sont dit "phylogénétiquement informatifs" lorsqu'ils sont partagés par au moins deux individus et au maximum tous les individus moins deux.

Outre les propriétés intrinsèques à la nature des caractères, l'analyse moléculaire présente l'avantage de pouvoir fournir un nombre de caractères quasi illimité. Tous les nucléotides constituant des génomes (nucléaire, chloroplastique, mitochondrial) sont des caractères potentiels. Ceci équivaut à dire que le riz, qui contient 1,2 pg d'ADN par noyaux (Arumuganathan & Earle, 1991 a), possède environ  $1,2 \cdot 10^9$  caractères potentiels sur le génome nucléaire (tout au long de la séquence nucléotidique). Cette caractéristique particulière conduit à des phylogénies présentant généralement moins de polytonies, et donc mieux résolues que les phylogénies basées sur des caractères conventionnels (Bremer & Struwe, 1992).

Deux techniques sont largement utilisées pour l'approche moléculaire des phylogénies : le séquençage qui est la seule méthode directe, et les RFLP qui mettent en évidence à la fois des variations de longueur de fragments de restriction (insertion/délétion) et des pertes ou gains de sites de restriction (substitution ou insertion/délétion).

### 2.1.2. Séquences

L'étude des séquences permet une comparaison directe, nucléotide après nucléotide de toute zone du génome : codante, transcrite (introns) ou non-codante et non-transcrite (zones intergéniques).



Les séquences de parties codantes peuvent être analysées directement au niveau de la séquence nucléique ou après traduction, sur la séquence protéique. Toutefois les analyses concomitantes sur la séquence du gène et de la protéine correspondante montrent que l'index de consistance (Farris, 1989) du cladogramme construit à partir des séquences nucléiques est meilleur (Bremer, 1988). L'intérêt phylogénétique des séquences d'acides aminés pâtit du manque de connaissance des mécanismes de l'évolution protéique et d'outil adapté à leur analyse.

Les alignement des séquences nucléiques mettent en évidence 3 types de mutations dites ponctuelles :

- les transitions, remplacement d'une purine (A,G) par une autre purine, ou d'une pyrimidine (C,T) par une autre pyrimidine ( $A \leftrightarrow G$  et  $C \leftrightarrow T$ );

- les transversions, que l'on classe en deux groupes, sont les remplacements d'une purine par une pyrimidine ou l'inverse. Les transversions de groupe 1 ne font pas intervenir de changement du nombre de liaisons hydrogènes ( $A \leftrightarrow T$  et  $C \leftrightarrow G$ ). Par contre, les transversions de groupe 2 impliquent une variation du nombre de liaisons hydrogènes, et donc de la stabilité ( $A \leftrightarrow C$  et  $G \leftrightarrow T$ )

- les indels (insertion/délétion) ; chaque indel, quelle que soit la longueur du fragment inséré ou délété, n'est considéré que comme un seul événement mutationnel.

Dans les zones codantes, les transitions et transversions peuvent conduire à un changement d'acide aminé ou à une terminaison de la traduction (codon stop) : dans le premier cas on parle de mutation faux-sens, dans le deuxième de mutation non-sens. Si la mutation sur la séquence nucléique n'entraîne pas de mutation sur la séquence protéique, la mutation est dite silencieuse.

L'interprétation phylogénétique des séquences nucléiques codantes et non-codantes est abordée différemment. Sur les séquences codantes, l'analyse peut être réalisée sur l'ensemble de la séquence ou seulement sur les premières, deuxième ou troisième positions des nucléotides dans les codons de la phase de lecture étudiée. Les taux d'évolution des "troisième positions" sont comparables à ceux des séquences non-codantes adjacentes. Les taux d'évolution des "premières" et "deuxièmes" positions des nucléotides dans les codons sont généralement plus faibles que ceux des séquences non-codantes adjacentes.

### 2.1.3. "RFLP"

L'analyse de type RFLP<sup>1</sup> produit indirectement des informations liées aux variations de séquences. Les mutations ponctuelles sont détectées par la perte ou le gain d'un site de restriction (présence de deux ou d'un fragments ayant pour taille la somme des deux précédents) ; les mutations en longueur sont repérées par des différences de taille des fragments.

Une même analyse de restriction (RFLP) peut être codée de façon différente. La méthode la plus simple est le codage direct des fragments de restriction. Bremer décrit deux façon de les coder (Bremer, 1991). La première fait intervenir tous les fragments révélant du polymorphisme, c'est à dire ceux phylogénétiquement informatifs (signe de synapomorphies) et ceux responsables des autapomorphies : c'est le codage dit FOA<sup>2</sup>. Le deuxième codage est dérivé du premier mais il ne prend en compte que les fragments phylogénétiquement informatifs (FDA<sup>3</sup>). Le codage des fragments (FOA ou FDA) a largement été utilisé associé au calcul d'une distance, suivie d'une agrégation par UPGMA ou Neighbor-Joining pour produire un dendrogramme (Schmit *et al.*, 1993; Sakata *et al.*, 1991; Pradhan *et al.*, 1992).

Afin de limiter le biais introduit par la redondance de certaines informations, il semble préférable lorsque cela est possible, de coder non plus les fragments mais les sites de restriction. Cela nécessite la reconstruction d'une carte de restriction plus ou moins détaillée. Ensuite, une analyse simple des autoradiographies permet de passer d'un type d'analyse à l'autre. La présence d'un fragment de 6kb à un endroit particulier pour un taxon, et de deux fragments de 2 et 4kb à la même position pour un autre taxon, peut être interprété comme la perte d'un site chez le premier ou le gain d'un site chez le second (Bremer, 1991). Une fois les sites identifiés, il existe également deux façons de les coder. La première (codage "SOA"<sup>4</sup>) utilise tous les sites polymorphes, quelle que soit l'origine du polymorphisme. Il peut être dû à une mutation sur l'une des 6 bases constituant le site, à une insertion ou à une délétion. Lorsque cela est possible on ne tient compte de la perte ou du gain d'un site que lorsque l'on suppose qu'il est dû à la même mutation (codage SMA<sup>5</sup>).

La majorité des travaux actuels utilisent le codage en sites (SOA ou SMA) associés à une analyse cladistique parcimonieuse (i.e. qui minimise le nombre de caractères sur les branches) pour la reconstruction d'arbres phylogénétiques (Bremer & Jansen, 1991; Jansen *et al.*, 1991;

<sup>1</sup>Restriction Fragment Length Polymorphism (polymorphisme de longueur de fragment de restriction)

<sup>2</sup>Fragment Occurrence Analysis

<sup>3</sup>Fragment Direct analysis

<sup>4</sup>Site Occurrence Analysis

<sup>5</sup>Site Mutation Analysis

Bremer & Struwe, 1992; Doebley *et al.*, 1992; Wilson *et al.*, 1992; Johansson & Jansen, 1993; Sobral *et al.*, 1994; Mayer & Soltis, 1994).

## 2.2. L'ANALYSE PHYLOGENETIQUE

L'arbre phylogénétique est le support formel de la représentation du processus dynamique de la diversification des espèces. Les espèces contemporaines y sont reliées par un réseau de branches (les liens de filiation) et de noeuds internes (les ancêtres virtuels). Différentes approches permettent de retrouver cet arbre à partir des données contemporaines. Deux types d'analyse, se rapportant à deux écoles de pensée différentes, coexistent :

- L'analyse cladistique inclue les procédures de parcimonie et les méthodes de compatibilité. Elle est directement basée sur une étude des caractères. C'est la méthode la plus utilisée pour l'interprétation phylogénétique du polymorphisme de sites de restriction (chaque site correspond à un caractère).

- L'analyse phénétique est basée sur l'étude des distances entre individus. L'arbre est construit à partir de la matrice de distance (espècesXespèces). C'est une approche largement employée pour la reconstruction phylogénétique à partir des données de séquences.

L'analyse cladistique est largement employée pour l'analyse des données moléculaires. Basée sur une analyse directe des caractères, elle autorise le "mélange" d'informations de différentes origines. Ceci signifie qu'elle permet d'analyser à la fois des données de type mutations ponctuelles et des insertions/délétions quelle que soit leur longueur, sans introduction d'un biais comme dans les méthodes de distance. L'analyse cladistique a été introduite par Hennig comme étant de la "systématique phylogénétique". Elle vise à identifier les états primitifs des caractères (dits "**plésiomorphes**") et les états dérivés (dits "**apomorphes**"). Les parentés entre les taxons étudiés sont identifiées sur la base des seuls états dérivés partagés par plusieurs taxons. Ces apomorphies partagées sont appelées "**synapomorphies**". Les synapomorphies sont imputées à l'héritage d'une espèce ancestrale propre aux taxons les possédant. On construit ainsi des groupes monophylétiques. Les apomorphies propres à un seul taxon terminal sont appelées "**autapomorphies**", elles ne permettent pas d'émettre d'hypothèse de parenté entre groupes. Si le taxon terminal est constitué d'un groupe d'espèces, les autapomorphies de ce taxon sont les synapomorphies des taxons qu'il regroupe. C'est une méthode empirique qui prétend ne faire

appel à aucun modèle sur les processus évolutifs. La seule hypothèse à vérifier est l'indépendance des caractères.

### 2.2.1. Notion de parcimonie

Toutes les procédures de parcimonie au sens cladistique, permettent de construire un arbre phylogénétique qui minimise l'homoplasie. L'homoplasie ou "bruit évolutif" est due à trois paramètres essentiels :

- la convergence, où deux individus A et C situés sur deux branches différentes évoluent tous les deux vers B,
- la réversion, où A évolue vers B qui lui même évolue vers A ( le passage par B pouvant passer inaperçu),
- le parallélisme, où deux individus A et B situés sur deux branches différentes vont acquérir la même mutation et évoluer respectivement vers A' et B'.

En pratique, pour utiliser les méthodes de parcimonie il faut étudier des zones de génome dans lesquelles le bruit évolutif est faible.

La recherche du sens de l'évolution des caractères de l'état primitif vers l'état dérivé revient à exprimer la similitude sous forme binaire 0-1. Cette approche est aisément informatisable, et plusieurs algorithmes de recherche d'arbre le plus court ont été proposés. Ces algorithmes sélectionnent l'arbre qui minimise la longueur totale des branches, c'est à dire celui pour lequel le nombre de "pas" nécessaire à la reconstruction de l'arbre est minimum.

Soit  $t$  l'arbre, il est choisi tel que  $L(t)$  soit minimum dans la formule :

$$L(t) = \sum_{k=1}^B \sum_{j=1}^N [W_j \text{diff}(x_{k'}^j, x_{k''}^j)]$$

où  $B$  est le nombre de branches,

$N$  est le nombre de caractères,

$k'$  et  $k''$  sont 2 noeuds incidents à chaque branche  $k$ ,

$\text{diff}(y,z)$  est une fonction spécifiant le coût de la transformation  $y$  en  $z$  le long de n'importe quelle branche,

et  $W_j$  assigne un poids à chaque caractère.

Dans la pratique, seuls les algorithmes dits "exacts" garantissent des solutions optimales. A partir de  $n$  taxons terminaux, on peut construire  $(2n-3)!/(2^{n-2} (n-2)!)$  arbres dichotomiques.

Pour 10 taxons terminaux, il existe donc 34 459 425 arbres possibles. La comparaison de tous ces arbres pour trouver le plus court peut vite devenir extrêmement laborieuse. La technique du "branch and bound" permet d'étudier des cas comprenant jusqu'à 20 ou 30 taxons, selon que les données comportent beaucoup ou peu d'homoplasie. Nous n'utiliserons pas ici cette famille de techniques qui ne peut être utilisée que dans des cas où la taille de l'échantillon est très limitée. Une alternative consiste à utiliser des algorithmes heuristiques. Ils ne garantissent pas la découverte de l'arbre optimal, mais permettent de traiter un plus grand nombre d'individus. Nous allons aborder un peu plus en détail la partie théorique de ce type d'analyse lorsqu'elle est appliquée à des données moléculaires binaires (0/1).

### 2.2.2. La parcimonie de Wagner

Dans ce modèle décrit par Kluge et Farris (1969), les convergences et réversions sont admises. C'est le modèle le plus simple puisqu'il n'impose que le minimum de contraintes pour le changement d'état des caractères.

Pour traiter cette hypothèse, Farris (1970) propose un algorithme qui permet d'établir des arbres non racinés.

-Soit  $Q_{xn}$ , l'état du caractère  $n$  chez l'individu  $x$  (dans notre hypothèse  $Q_{xn}$  est donc égal à 0 ou 1).

-La différence entre les deux individus  $x$  et  $y$  est définie par la distance de Manhattan établie à partir des  $N$  caractères :

$$d_{xy} = \sum_{n=1}^N |Q_{xn} - Q_{yn}|$$

Cette différence représente exactement le nombre de pas nécessaires pour relier  $x$  à  $y$ .

- $x$  et  $y$  pour lesquels cette distance est maximale, sont connectés.

-Un troisième individu  $z$  est connecté à un point entre  $x$  et  $y$  (premier noeud appelé  $k$ ) tel que la distance entre  $z$  et  $k$  soit maximale.

-Pour l'insertion suivante un individu  $w$  est de même connecté à un noeud secondaire ( $k'$ ) choisi tel que la distance entre ce noeud et  $w$  soit maximale. Ce processus est ainsi répété jusqu'à ce que tous les individus soient agglomérés (Farris, 1970).

Cette méthode heuristique ne conduit pas forcément à l'arbre de longueur minimale. Il est nécessaire d'optimiser sa topologie, soit en changeant le point de départ de l'agglomération, soit en effectuant des réarrangements de branches.

### 2.2.3. Les autres parcimonies

Dans les arbres (ou réseaux) de Wagner le postulat du critère de parcimonie s'énonce comme suit : dans une unité de temps donnée la probabilité d'avoir la mutation de 0 en 1 est la même que celle d'avoir 1 en 0.

Dans le modèle dit de Dollo, les convergences sont exclues. Seules sont admises les réversions. Il y est postulé qu'en cas d'homoplasie, il est plus facile de perdre un caractère par réversion que d'acquérir le même caractère sur une branche parallèle.

A l'inverse il est postulé dans le modèle de Camin-Sokal que l'évolution est irréversible (les convergences sont admises mais les réversions sont interdites), ce qui pour des informations de type moléculaire est fortement improbable.

## 2.3. QUEL GÉNOME ETUDIER ?

L'essor de la biologie moléculaire au cours de ces dernières années, l'amélioration des techniques d'isolement de l'ADN, la généralisation de la PCR<sup>6</sup> et le séquençage "en routine" ont permis d'améliorer considérablement notre connaissance des génomes des plantes. Les travaux concernant la phylogénie sont basés sur les génomes cytoplasmiques (chloroplastique et mitochondrial) et nucléaires (gènes "simple copie" et gènes codants pour les ARN ribosomiques). Du point de vue de la reconstruction de phylogénie, les potentialités d'un génome ou d'une région de ce génome dépendent d'un certain nombre de facteurs extérieurs et intrinsèques : abondance, accessibilité, taille, mode d'évolution et niveau d'homoplasie. Le choix du génome à étudier est également dépendant du niveau taxonomique considéré.

---

<sup>6</sup>Polymerase Chain Reaction (réaction de polymérisation en chaîne)

### 2.3.1. Génome mitochondrial

L'ADN mitochondrial (ADNmt) a été largement utilisé pour étudier la phylogénie chez les animaux. En effet l'ADNmt animal est très conservé en taille (16 kb en moyenne chez les métazoaires) et en structure. La divergence intragénérique des séquences est variable de 5 à 10% suivant les gènes et les régions. Les positions relatives des gènes varient peu chez les vertébrés. On ne trouve des différences qu'entre grands groupes : vertébrés, insectes, échinodermes. Les séquences d'ADNmt d'animaux évoluent à un taux très élevé : le taux de substitutions est 10 à 30 fois plus élevé que dans les séquences d'ADN chloroplastique chez les plantes (Palmer, 1992). De plus le rapport des fréquences transitions/transversions est toujours très élevé, de l'ordre de 20 (Zurawski & Clegg, 1987).

La situation des ADNmt de plantes est complètement différente. Les ADN mitochondriaux végétaux sont de taille plus élevée et extrêmement plus variable (217 kb chez *Brassica sp.*, 570 kb chez *Zea mays*) que les ADNmt de champignons ou d'animaux (15 à 16kb chez les mammifères) (Leaver, 1988). Dans deux familles (Cucurbitaceae et Malvaceae) on a rapporté des tailles dépassant 2500kb (Palmer, 1992). La recombinaison homologue entre courtes séquences répétées y est très fréquente. Elle conduit à la présence dans une même plante d'une population d'ADNmt de taille hétérogène (Leaver, 1988). Paradoxalement alors que l'ADNmt réarrange très rapidement par inversions et recombinaisons, les gènes mitochondriaux végétaux qui ont été séquencés semblent évoluer très lentement. En effet, le taux de substitution est environ 40 à 100 fois plus faible pour les gènes mitochondriaux végétaux que pour les gènes mitochondriaux animaux (Palmer *et al.*, 1988; Palmer, 1992). Le taux de substitution des gènes mitochondriaux végétaux est également plus faible (3 à 4 fois) que celui des gènes chloroplastiques.

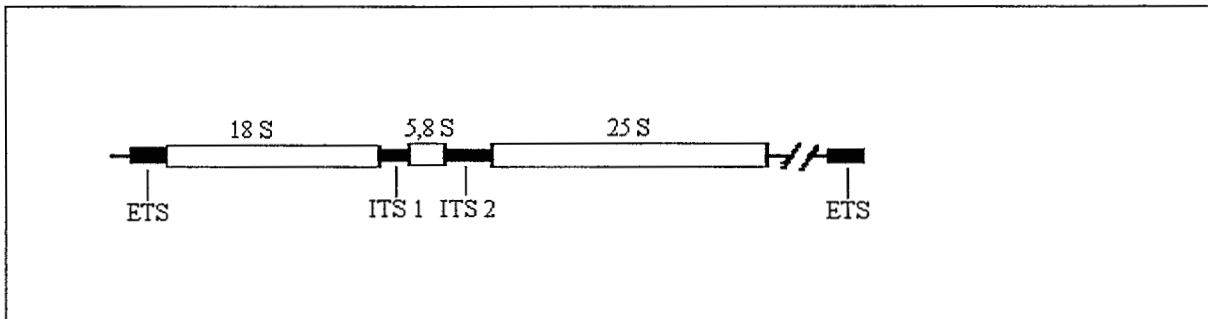
Ces caractéristiques font du génome mitochondrial végétal un matériel difficile à étudier par l'analyse des fragments de restriction. De plus, étant donné le très faible taux d'évolution par mutations ponctuelles, le séquençage de ce génome ne présente d'intérêt que pour la reconstruction phylogénétique à très haut niveau taxonomique (Palmer, 1992).

### 2.3.2. Gènes codants pour les ARNs ribosomiques

Ces gènes sont présents dans les trois génomes des plantes : nucléaire, chloroplastique et mitochondrial. Les différentes régions constitutives des ARN ribosomiques (ARNr) n'évoluent pas à la même vitesse, ce qui les rend utilisables à plusieurs niveaux taxonomiques, de

l'intraspécifique jusqu'à l'inter-générique. Ce sont essentiellement les séquences nucléaires qui ont été utilisées en phylogénie.

Les gènes nucléaires codant pour les ARNr sont organisés en tandem de plusieurs centaines de répétitions d'une unité appelée "unité d'ADNr" (situés sur un ou plusieurs loci chromosomiques). Ces unités d'ADNr sont séparées entre elles par une zone inter-génique non-codante ou "IGS" (intergenic spacer). Chaque unité d'ADNr est constituée de trois parties codantes et d'espaces (figure 2). L'unité transcrite (ETS, 18S, ITS1, 5,8S, ITS2, 25S) a une taille presque constante de 6 kb. Au contraire, l'espaceur intergénique a une taille très variable, de 1 à 8 kb, qui induit une variation de taille du simple au double de l'unité d'ADN ribosomique (Hamby & Zimmer, 1992). Il existe une autre unité d'ADNr, codant pour l'ARN ribosomique 5S également répétée en tandem et dont les copies sont séparées par un espaceur non transcrit (NTS).



**Figure 2.** Représentation schématique d'une unité d'ADNr. Les zones en noir figurent les régions transcrites et non traduites, externes (ETS) et internes (ITS).

Dans l'unité "type" les parties codantes sont hautement conservées, alors que les espaces qu'ils soient ou non transcrits sont sujets à des variations.

Les travaux concernant les variations de taille de l'IGS ont permis de mettre en évidence des différences intra spécifiques (Saghai-Maroo *et al.*, 1984; May & Appels, 1987; Appels & Dvorak, 1982). Ces travaux ont produit des marqueurs génétiques informatifs pour l'étude des populations, et des variations entre individus d'une même espèce (Sano & Sano, 1990; King & Schaal, 1989; Harding, 1991; Lanaud *et al.*, 1991). Ils sont basés sur des analyses de type RFLP. Le haut niveau de polymorphisme peut créer des situations complexes qui dans tout les cas nécessitent la construction d'une carte de restriction pour que l'analyse soit correcte (Doebley, 1992).



Le séquençage des régions codantes met en évidence un très faible niveau de variation (Schaal & Learn, 1988). Il est donc utilisé à très haut niveau taxonomique entre angiospermes et gymnospermes, monocotylédones et dicotylédones (Hamby & Zimmer, 1992). Du point de vue phylogénétique, l'étude de ces séquences montre une origine unique des plantes à fleurs, au sein duquel les angiospermes constituent un groupe monophylétique (Hamby & Zimmer, 1992).

Plus récemment le séquençage de régions transcrites non-codantes (ITS), a permis d'étudier des relations phylogénétiques à un niveau inférieur (Suh *et al.*, 1993; Hsiao *et al.*, 1994).

Le type d'approche développée sur l'IGS l'est également sur le NTS (fragment intergénique des ADN<sub>r</sub> codant pour l'ARN-5S) qui, par sa petite taille (inférieure à 1 kb) peut être directement amplifiée par PCR : la "technique" est plus facile, mais l'analyse des profils reste assez complexe. L'interprétation rigoureuse du polymorphisme nécessite le séquençage (Cox *et al.*, 1992).

### 2.3.3. ADN chloroplastique

L'ADN chloroplastique (ADN<sub>cp</sub>) est une molécule circulaire bicaténaire située à l'intérieur des chloroplastes, dans le cytoplasme des végétaux supérieurs. Le nombre de chloroplastes par cellule (de l'ordre de 30 à 500), tout comme le nombre de molécules d'ADN<sub>cp</sub> par chloroplaste (entre 20 et 200) sont extrêmement variables (Palmer, 1987). On estime à quelques milliers le nombre de copies d'ADN<sub>cp</sub> identiques dans chaque cellule végétale d'un individu (distribution de type uniclonal).

Deux types de mutations coexistent chez l'ADN<sub>cp</sub> : les mutations ponctuelles et les réarrangements (Palmer *et al.*, 1988) :

- Les mutations les plus fréquentes sont les mutations ponctuelles (transitions, transversions) et les insertions/délétions (indels) courtes dans les régions non-codantes de l'ADN<sub>cp</sub>. Les mutations ponctuelles sont largement utilisées pour les reconstructions phylogénétiques. Elles ont l'intérêt de présenter très peu d'homoplasie, ce qui garantit une reconstruction plus robuste à partir de données chloroplastiques plutôt qu'à partir de données provenant d'autres génomes. Les indels sont essentiellement retrouvés dans les régions non-codantes où ils sont organisés de façon contiguë. Lors d'une analyse de type RFLP, ils sont difficilement codables. Sachant que ces erreurs de codage peuvent induire de l'homoplasie préjudiciable à la robustesse de l'arbre

phylogénétique (Palmer *et al.*, 1988), les indels repérés par RFLP ne sont généralement pas utilisés comme des caractères phylogénétiquement informatifs.

- Les réarrangements sont des événements beaucoup plus rares au cours de l'évolution de l'ADNcp (Palmer *et al.*, 1988). Ce sont des inversions ou des insertions/délétions de gènes entiers, d'introns, parfois même de plusieurs gènes ou d'une copie de l'IR. Leur rareté font de ces changements des éléments souvent très utiles pour supporter des embranchements phylogénétiques à haut niveau taxonomique (Jansen *et al.*, 1991; Johansson & Jansen, 1993).

En fonction du niveau phylogénétique et de la différenciation des espèces étudiées, plusieurs types d'approches moléculaires sont envisageables pour mettre en évidence les variations de l'ADNcp.

#### 2.3.4. Avantages comparatifs

Les recombinaisons de l'ADNmt rendent son analyse par la technique des RFLP difficilement exploitable. L'analyse des séquences s'impose alors le plus souvent. Chez l'ADNmt le taux de substitution très faible le rend utilisable seulement pour l'étude des relations de parenté à très haut niveau taxonomique. Par contre l'étude des gènes et des zones intergéniques du génome nucléaire semble capable de fournir des informations phylogénétiques sur des unités taxonomiques proches. Bien que les gènes en simple copie (ou faible nombre de copies) puissent fournir des résultats cohérents (Gaut & Clegg, 1991), ce sont le plus souvent les gènes et inter-gènes (Suh *et al.*, 1993; Hsiao *et al.*, 1994) des régions codant pour les ARN ribosomiques qui sont utilisés. Cette approche se révèle puissante à tous les niveaux phylogénétiques et ceci jusqu'au niveau intraspécifique (Faivre-Rampant *et al.*, 1992; Schaal & Learn, 1988). Elle fournit des informations sur l'ascendance paternelle et maternelle (Hamby & Zimmer, 1992).

L'ADNcp demeure le plus largement utilisé pour les études phylogénétiques chez les végétaux, et ce, quel que soit le niveau taxonomique : il possède un mode d'évolution conservateur et présente l'avantage, malgré un taux d'évolution relativement faible (comparativement aux séquences nucléaires) de fournir des caractères exceptionnellement peu homoplasiques pour les études phylogénétiques. Son hérédité le plus souvent monoparentale, rend la première approche de la phylogénie plus facile à analyser.

## 2.4 ANALYSE DES VARIATIONS DE L'ADNCP

### 2.4.1. Structure de l'ADNcp

Chez les angiospermes la molécule circulaire d'ADNcp a une taille comprise entre 120 à 217 kb (Palmer, 1985). Cette taille est beaucoup plus variable pour les algues vertes chez qui elle peut atteindre des valeurs supérieures à 1000 kb avec une forme ADN bicaténaire linéaire chez *Acetabularia*. Parmi les plantes supérieures, ce sont les légumineuses qui ont le plus petit génome chloroplastique (120 à 140 kb).

La plus grande partie des plantes supérieures possède une molécule d'ADNcp divisée en quatre régions (Palmer, 1985). La carte de l'ADNcp de riz, qui peut servir de modèle pour la majorité des plantes supérieures, est présentée sur la figure 3 :

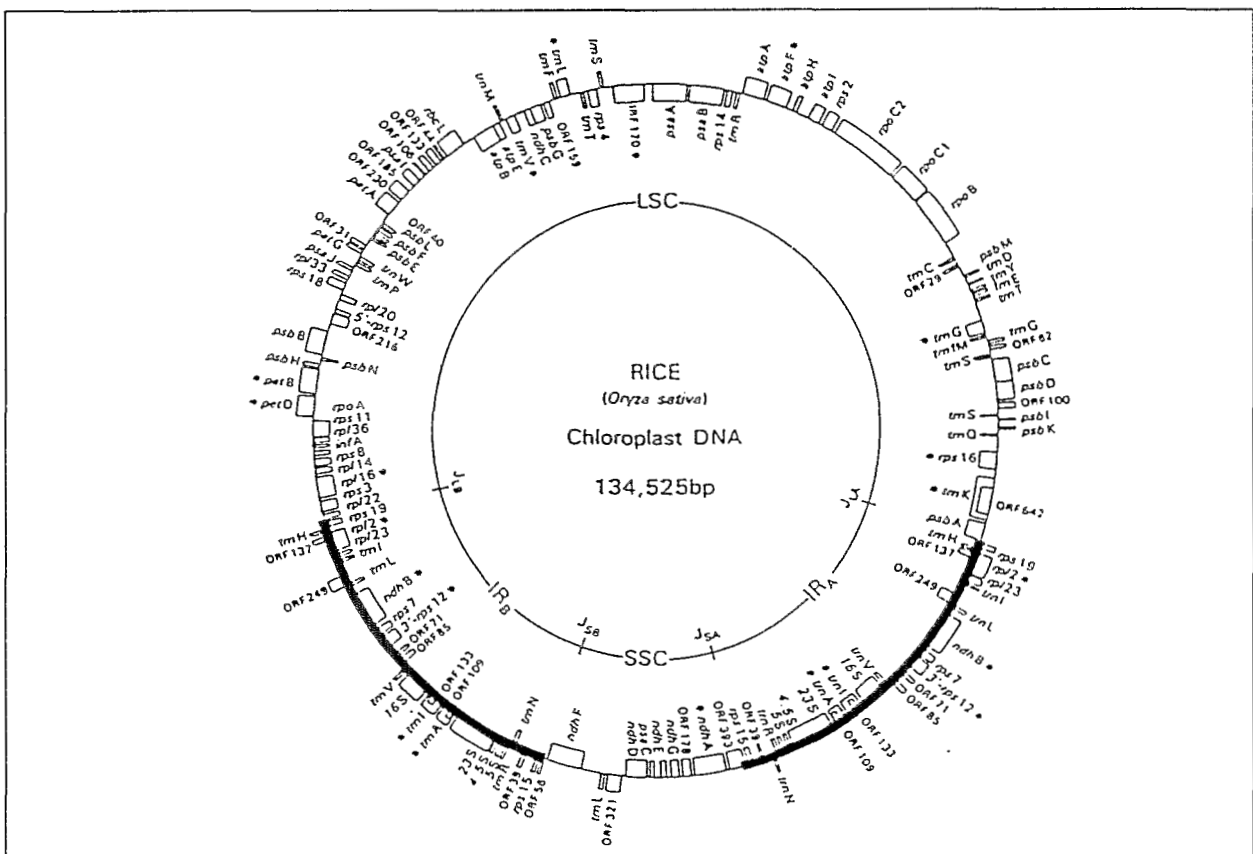


Figure 3. Carte de l'ADNcp d'*Oryza sativa* (Hiratsuka *et al.*, 1989)

Un grand fragment (21 kb chez *Oryza sativa*), dupliqué en orientation opposée, forme la région inversée répétée (Inverted-Repeat Région, ou IR). Cette région soumise aux

recombinaisons, est responsable de la présence du mélange équimolaire de deux isomères génétiquement identiques mais physiquement différents. Il semble que la perte d'un des deux segments de l'IR soit très souvent associée à un taux élevé de réarrangements. Cette région est également la principale responsable des variations de la taille du génome chloroplastique. La faible taille de l'ADNcp des légumineuses fait suite à la perte d'une copie entière d'une IR. A l'inverse, les plus gros génomes (217kb chez le géranium) sont liés à l'augmentation de la taille des régions inversées répétées.

Le reste de la molécule comprend deux régions en copie unique, une petite appelée petite région simple copie ou Small Single Copy (SSC) et une plus grande ou Large Single Copy (LSC). Elles mesurent respectivement environ 12 et 80 kb chez le riz.

Outre la région IR, le génome chloroplastique possède également plusieurs types de séquences répétées : les répétitions en tandem, les répétitions dispersées et de nombreuses séquences répétées d'une longueur inférieure à 200pb. Le premier type de répétition (en tandem) a été décrit chez l'algue *Acetabularia mediterranea* et également chez *Euglena gracilis* (Palmer, 1985). Ce sont des répétitions de plusieurs kilobases en quatre ou cinq exemplaires.

Hormis les variations de taille liées au polymorphisme des IR, il existe d'autres mutations qui modifient la taille du génome chloroplastique par insertion ou délétion de fragments de séquences nucléotidiques. Elles sont de deux types : des mutations portant sur des fragments de 50 à 1000 pb qui sont assez rares, et des mutations beaucoup plus fréquentes principalement localisées dans les régions non-codantes et portant sur une dizaine de paires de bases (Palmer, 1985).

### 2.4.2. Organisation des gènes

A l'heure actuelle seuls les génomes chloroplastiques de trois plantes, une hépatique à thalle, *Marchantia polymorpha* (121kb) (Ohyama *et al.*, 1986), le riz (135 kb) (Hiratsuka *et al.*, 1989) et le tabac (156 kb) (Shinozaki *et al.*, 1986) ont été entièrement séquencés.

Il apparaît que le nombre de gènes diffère peu suivant les espèces. Il est d'environ 120 gènes différents pour les plantes terrestres (algues exclues). Ces gènes codent pour 4 ARNs ribosomiques, 30 ou 31 ARNs de transfert et 85 protéines intervenant principalement dans la photosynthèse et le système de transcription/traduction (Palmer *et al.*, 1988). Il semble que la perte ou le gain d'un gène constitue un événement extrêmement rare. En outre il est plus

approprié de parler de transfert puisque des gènes chloroplastiques ont été retrouvés dans le génome nucléaire ou dans le génome mitochondrial (Palmer, 1985). C'est le cas du gène *tufA* qui code pour le facteur d'élongation EF-Tu de la synthèse des protéines chloroplastiques, et qui a été transféré du chloroplaste au noyau il y a environ 400 millions d'années chez l'ancêtre commun aux plantes terrestres (Baldauf & Palmer, 1990).

Les gènes codant pour les ARNs ribosomiques sont généralement situés dans la région IR (Palmer, 1985) et donc présents en deux copies de même ordonnancement.

Les gènes codant pour les ARNs de transfert (ou *trn*) sont généralement situés dans les régions simples copies (SSC ou LSC). C'est le cas de 23 *trn* chez le tabac et de 26 chez le riz (Maréchal-Drouard *et al.*, 1993). Ils sont distribués sur l'ensemble de l'ADNcp, mais parfois organisés en groupes ou clusters séparés par de courts fragments intergéniques (Maréchal-Drouard *et al.*, 1993). Au moins 6 gènes codant pour des ARNt sont interrompus par des introns. Il s'agit du *trnA* (gène codant pour l'ARNt de l'alanine), *trnG* (glycine), *trnL* (leucine), *trnI* (isoleucine), *trnK* (lysine) et du *trnV* (valine) (Maréchal-Drouard *et al.*, 1993). Ces introns sont très longs (325 à 2526pb) comparés à la taille des ARNt matures. Ils sont épissés lors de la maturation.

### 2.4.3. Transmission

Chez les plantes supérieures l'ADNcp semble transmis (transmission clonale) le plus souvent de façon uni-parentale. Il est transmis par le parent femelle chez la plupart des angiospermes, et par le parent mâle chez les gymnospermes (Reboud & Zeyl, 1994).

L'hérédité cytoplasmique a été montrée par RFLP, par cytogénétique ou encore par coloration après traitement à la mitramicine (Reboud & Zeyl, 1994). Dans certains cas on observe un héritage biparental de l'ADNcp (Masoud *et al.*, 1990). Selon Palmer il n'existe aucun cas de fusion ou de recombinaison entre les deux molécules d'ADNcp chez les plantes présentant ce mode d'héritage (Palmer *et al.*, 1988). Pourtant des travaux sur *Pinus contorta* et *P. banksiana*, montrent que, dans certains cas, des profils pour les plantes hybrides diffèrent des profils observés pour les deux parents. Ces profils "polymorphes" résulteraient de contaminations, de chimérisme parental, de transfert maternel ou de recombinaisons (Wagner *et al.*, 1987; Govindaraju *et al.*, 1988).

#### 2.4.4. Analyse des sites de restrictions de l'ADNcp

Deux méthodes ont été envisagées pour étudier les variations de site ou de longueur de fragments de restriction (RFLP) sur le génome chloroplastique.

Les premières études étaient basées sur l'isolement de l'ADNcp, sa digestion par des enzymes de restriction puis la visualisation directe des bandes après migration et coloration du gel (Berthou *et al.*, 1983; Enomoto *et al.*, 1985). En fait, ce type d'analyse est limité aux situations pour lesquelles l'interprétation des profils en terme de mutations est possible (Palmer *et al.*, 1988) ; c'est à dire lorsque l'ADNcp est isolé en quantité suffisante (Dally & Second, 1990) et que la divergence des séquences ne dépasse pas 0,5 ou 1% (Pradhan *et al.*, 1992; Sakata *et al.*, 1991; Schmit *et al.*, 1993; Hooglander *et al.*, 1993).

Actuellement la majorité des travaux de type RFLP sur l'ADNcp est réalisée directement à partir d'ADN total digéré, hybridé à des sondes d'ADNcp homologues ou hétérologues après migration sur gel et transfert sur membrane (ceci est possible grâce à la grande conservation des séquences et de l'ordre des gènes de l'ADNcp). Non seulement cette approche requiert moins d'ADN (signal amplifié par l'utilisation de sondes "marquées") et simplement de l'ADN total, mais elle permet aussi la cartographie directe des sites de restriction et la mise en évidence des mutations de longueur de fragments. Elle se révèle donc plus générale et plus puissante (en terme d'analyse) que la précédente.

L'approche RFLP a été largement employée pour les reconstructions phylogénétiques entre espèces proches (même genre ou genres voisins). Elle peut également être utile au niveau intraspécifique (Terauchi *et al.*, 1991). Elle permet de mettre en évidence à la fois des mutations ponctuelles, des indels et des réarrangements. Généralement, l'approche phylogénétique utilise les mutations ponctuelles et le codage SOA ou SMA décrit au paragraphe 2.1.3.

Le niveau d'homoplasie mis en évidence est habituellement faible mais peut atteindre des valeurs plus élevées, jusqu'à 15% (Doyle *et al.*, 1990). De façon générale, la divergence nucléotidique au niveau intragénérique ne dépasse pas les 0,2% (Schilling *et al.*, 1994; Crawford *et al.*, 1990). Elle peut néanmoins être beaucoup plus faible, comme dans le cas des lentilles, le genre *Lens* (Mayer & Soltis, 1994), ou beaucoup plus élevée, comme dans le genre *Plantago* (Wolff & Schaal, 1992). Cette variation peut être liée au temps de divergence du genre, mais également à des taux d'évolution variables suivant les branches phylogénétiques de l'arbre. La robustesse de la phylogénie est dépendante du nombre de caractères

phylogénétiquement informatifs pris en compte. Ce nombre dépend à la fois du nombre de sites étudiés (liés à l'expérimentation) et des deux paramètres précédemment cités (temps de divergence, vitesse d'évolution). (Systma & Gottlieb, 1986; Doebley *et al.*, 1992; Wilson *et al.*, 1992; Choo *et al.*, 1994; Duvall *et al.*, 1994; Gielly & Taberlet, 1994)

#### 2.4.5. Séquences

Palmer précise que la divergence nucléotidique sur les parties codantes de l'ADNcp entre espèces congénériques est habituellement de l'ordre de 2% (Dowling *et al.*, 1990). Toutefois ce taux est extrêmement variable d'un gène à un autre, d'un groupe taxonomique à un autre.

Le gène chloroplastique *rbcL* (codant pour la grande sous-unité de la ribulose-1,5-bisphosphate carboxylase/oxygénase ou RuBisCO) a été sans conteste le plus largement utilisé. Son étude est particulièrement intéressante pour les phylogénies à haut niveau taxonomique et ses 1428 paires de bases (1458 pour les Asteraceae) ont été séquencées chez près de 500 espèces différentes, représentant 265 familles de plantes supérieures (Chase *et al.*, 1993). Ce gène a permis notamment d'étudier les relations de parenté chez les végétaux à l'intérieur d'une sous-classe (Magnoliidae (Qiu *et al.*, 1993), Asteridae (Olmstead *et al.*, 1993)) et entre familles (chez les monocotylédones (Duvall *et al.*, 1993), entre Ericaceae, Empetraceae et Epacridaceae (Kron & Chase, 1993)). Les résultats obtenus démontrent que des phylogénies sûres peuvent être résolues à ces niveaux taxonomiques élevés. Les relations phylogénétiques ont été également étudiées à l'aide des données de séquence du gène *rbcL* à des niveaux taxonomiques plus bas (inter et intra génériques), par exemple chez les Cupressaceae (Gadek & Quinn, 1993), les Geraniaceae (Price & Palmer, 1993), les Cornaceae (Xiang *et al.*, 1993), les Asteraceae (Michaels *et al.*, 1993), les Datisceae (Swensen & Mullin, 1994), indiquant que *rbcL* peut être employé jusqu'au niveau du genre. Toutefois, dans certains cas, de nombreuses relations demeurent imprécises, comme chez les Cornaceae (Xiang *et al.*, 1993) ou chez les Saxifragaceae (Soltis *et al.*, 1993). Le gène *rbcL* semble souvent trop conservé pour éclaircir les relations entre genres proches, et à l'intérieur d'un même genre.

Plus récemment, un certain nombre de travaux ont débuté sur le gène *matK* (Steele & Vilgalys, 1994; Johnson & Soltis, 1994). Le travail réalisé chez les Saxifragaceae montre une très bonne cohérence des phylogénies obtenues à partir des données de séquences de *matK* et des données RFLP sur l'ensemble du génome chloroplastique (Johnson & Soltis, 1994). Dans ce cas précis, le séquençage du gène *matK* a permis de descendre en dessous du pouvoir de résolution du gène *rbcL*.

L'analyse des régions non-codantes devrait étendre l'utilité de la molécule d'ADN chloroplastique aux niveaux taxonomiques encore plus bas (Gielly & Taberlet, 1994). Ces régions ont en effet tendance à évoluer plus rapidement que les régions codantes, en partie par accumulation d'insertions et de délétions (Zurawski & Clegg, 1987; Clegg & Zurawski, 1992; Clegg, 1993) et peuvent donc devenir très utiles en dessous du niveau de la famille (Gielly, 1994). Au niveau intra générique, l'étude de séquences non-codantes doit permettre la construction de phylogénies mieux résolues (Gielly & Taberlet, 1994). Actuellement deux types de zones non-codantes de l'ADNcp ont été étudiées : la région inter-génique *atpB-rbcL* (Savolainen *et al.*, 1993; Spichiger *et al.*, 1993; Savolainen *et al.*, 1994) et trois zones adjacentes aux gènes codants pour les ARN de transfert ou *trn*, *trnL*, *trnF* et *trnT* (Gielly & Taberlet, 1994).



## **2.5 BILAN**

Les études basées sur une approche moléculaire produisent généralement des phylogénies bien résolues du fait du nombre élevé de caractères disponibles. Par ailleurs, les marqueurs produits par ce type d'analyse sont de bons marqueurs génétiques : ceci signifie que les phylogénies établies à partir de caractères moléculaires sont probablement proches de la réalité biologique.

En fonction des individus dont on veut étudier les relations phylogénétiques, il est nécessaire de choisir judicieusement la ou les zones de génome à analyser pour obtenir un niveau maximum de résolution.

Le choix du génome à analyser, dépend du matériel à étudier. Chez les plantes supérieures l'ADNcp constitue un support privilégié puisqu'il produit généralement des reconstructions phylogénétiques comportant peu d'homoplasie.

Le choix des zones du génome à étudier est fonction du niveau taxonomique auquel on se place. L'étude de régions conservées est réservée aux niveaux taxonomiques élevés. Pour l'étude des relations phylogénétiques entre genres proches ou entre espèces d'un même genre les séquences codantes ne permettent pas de construire des phylogénies bien résolues. A ces bas niveaux taxonomiques, c'est l'étude de régions moins conservées qui est la plus informative.

Enfin, le choix de la méthode d'analyse est directement dépendant du génome et/ou de la zone que l'on a choisi d'étudier.

### **3. VARIATIONS DE L'ADN CHLOROPLASTIQUE DES CAFEIERS**

L'étude de la diversité moléculaire, particulièrement celle se basant sur l'ADNcp, constitue un outil puissant pour l'évaluation des relations phylogénétiques. C'est pourquoi nous avons choisi cette voie pour tenter d'établir les relations de parenté entre les espèces de caféiers. Cette étude a pour but de dégager des hypothèses concernant les voies de spéciation, mais également de préciser la place et le rôle phylogénétique d'espèces jamais étudiées auparavant.

Dans un premier temps nous avons étudié le polymorphisme de l'ADNcp dans son ensemble, par RFLP. Les résultats obtenus nous ont amenés à poursuivre cette approche par une étude plus spécifique de régions non-codantes de l'ADNcp. Nous allons présenter séparément les résultats et les informations obtenues par ces deux approches. Dans un troisième temps, une analyse phylogénétique de l'ensemble des résultats sera menée (chapitre 5).

L'objectif principal de cette première étude était d'estimer le polymorphisme de l'ADNcp chez les caféiers. Il devait permettre également de préciser la structure de l'ADNcp des caféiers et de la comparer à celle d'autres espèces.

L'étude du génome chloroplastique d'une espèce ou d'un groupe d'espèces peut, grâce aux propriétés de ce génome (conservé, abondant, évoluant principalement par mutations ponctuelles), être facilement entreprise à l'aide de sondes hétérologues. Bremer et Jansen, dans leur travail sur un grand nombre d'espèces de *Rubiaceae*, ont utilisé des sondes provenant de l'ADNcp de la laitue (Bremer & Jansen, 1991). L'appartenance taxonomique des caféiers à la famille des *Rubiaceae* nous a incité à travailler avec le même jeu de sondes.

### 3.1 MATERIEL

#### 3.1.1. Matériel Végétal

Le matériel végétal a été choisi de façon à représenter le mieux possible la diversité génétique disponible. Malgré nos efforts, une espèce malgache (*C. sakaraha*), un échantillon du genre *Coffea* sous-genre *Baracoffea* (*C. grevei*) et un représentant du genre *Psilanthus* sous-genre *Afrocoffea* (*P. bengalensis*), n'ont pu être étudiés faute d'avoir pu obtenir des feuilles en quantité suffisante et/ou de l'ADN en quantité et qualité correcte (non dégradé). Le matériel que nous avons étudié "recouvre" néanmoins les trois zones biogéographiques précédemment décrites. Les espèces africaines du sous-genre *Coffea* actuellement disponibles en collection en Côte d'Ivoire sont toutes représentées.

# Variation de l'ADN chloroplastique... / Chapitre 3

Espèce	Nom de génotype	Population sauvage/Variété cultivée	pays d'origine	n°
<i>C.arabica</i>	ET 12-5	sauvage, pop. Limu (core collection)	Ethiopie	211
<i>C.arabica</i>	Catura amarillo	variété cultivée		236
<i>C.arabica</i>	3099	sauvage, pop. Marsabit	Kenya	278
<i>C.arabica</i>	3058	sauvage, pop. Marsabit	Kenya	282
	HDT 832-1	descendant de l'hybride de Tymor		325
<i>C.bertrandii</i>		sauvage	Madagascar	408
<i>C.brevipes</i>	Mont Cameroun	sauvage, pop. Mont Cameroun	Cameroun	267
<i>C.brevipes</i>	Kumba-Loum	sauvage, pop. Mungo	Cameroun	276
<i>C.canephora</i>	IF 444	clone cultivé de type Congolais		264
<i>C.canephora</i>	IF 200	clone cultivé de type hybride Congolais/Guinéen		265
<i>C.canephora</i>	IF 155	clone cultivé de type Guinéen		279
<i>C.canephora</i>	IF A25	clone cultivé de type Congolais		284
<i>C.canephora</i>	IF 410	clone cultivé de type Guinéen		285
<i>C.canephora</i>	160 02	descendant d'1 clone cultivé hybride Congolais/Guinéen		288
<i>C.congensis</i>	03.255	sauvage, pop. Louma	Centrafrique	261
<i>C.congensis</i>	03.1650	sauvage, pop. Brazzaville 1	Congo	290
<i>C.congensis</i>	03.429	sauvage, pop. Anginga	Centrafrique	291
<i>C.congensis</i>	03.103	sauvage, pop. M'Boulou	Centrafrique	298
<i>C.costatifructa</i>	08.111	sauvage, pop. Ute 1	Tanzanie	302
<i>C.eugenioides</i>	04.1485	sauvage, pop. Cheptuyet	Kenya	263
<i>C.eugenioides</i>	04.010	sauvage, pop. Malava	Kenya	293
<i>C.eugenioides</i>	04.005	sauvage, pop. Kimilili	Kenya	294
<i>C.farafanganensis</i>		sauvage	Madagascar	329
<i>C.humboldtiana</i>	OB.080	sauvage	Comores	330
<i>C.humilis</i>	07.141	sauvage, pop. Sakré	Côte d'Ivoire	258
<i>C.kapakata</i>		introduction en provenance du Bresil	Angola	270
<i>C.kapakata</i>		introduction en provenance de Tanzanie	Angola	271
<i>C.liberica</i>	EC5	sauvage, pop. Koto (core collection)	Cameroun	303
<i>C.liberica (dewrevrei)</i>	05.797	sauvage, pop. N'Dongue	Centrafrique	257
<i>C.liberica (dewrevrei)</i>	05.559	sauvage, pop. Balifondo	Centrafrique	280
<i>C.liberica (liberica)</i>	EA1	sauvage (core collection)	Côte d'Ivoire	217
<i>C.liberica (liberica)</i>	05.242	sauvage, pop. Taï	Côte d'Ivoire	260
<i>C.millotii</i>		sauvage	Madagascar	308
<i>C.perrieri</i>		sauvage	Madagascar	268
<i>C.pervilleana</i>		sauvage	Madagascar	238
<i>C.pseudozanguebariae</i>	08.228	sauvage, pop. Uzigua	Tanzanie	254
<i>C.pseudozanguebariae</i>	08.021	sauvage, pop. Shimba	Kenya	255
<i>C.racemosa</i>		introduction en provenance de Tanzanie	Mozambique	209
<i>C.racemosa</i>		introduction en provenance du Portugal	Mozambique	253
<i>C.salvatrix</i>		introduction en provenance de Tanzanie	Bostwana	250
<i>C.salvatrix</i>		introduction en provenance du Bresil	Mozambique	251
<i>C.sessiliflora</i>	PA4	sauvage, pop. Shimba (core collection)	Kenya	225
<i>C.sessiliflora</i>	08.161	sauvage, pop. Kitulangalo	Tanzanie	313
<i>C.sp Mayombe</i>		sauvage, pop. Mayombe	Congo	324
<i>C.sp Mouloundou</i>	OC.210	sauvage, pop. Souanké	Congo	256
<i>C.sp N'gongo 2</i>	OC.282	sauvage, pop. Congo (Cg) 45	Congo	259
<i>C.sp N'koumbala</i>	OC.105	sauvage, pop. Cameroun (C) 05	Cameroun	275
<i>C.sp X</i>		sauvage, affine de <i>C.brevipes</i> variété heterocalyx		243
<i>C.stenophylla</i>	FB1	sauvage, pop. Ira (core collection)	Côte d'Ivoire	226
<i>C.stenophylla</i>	FA 21	sauvage, pop. Assabli (core collection)	Côte d'Ivoire	232
<i>P.ebracteolatus.</i>	OA.153	sauvage, pop. Bafingdala (core collection)	Côte d'Ivoire	248
<i>P.mannii</i>	OA.009	sauvage, pop. Divo (core collection)	Côte d'Ivoire	247

**Tableau 7.** Matériel végétal étudié. L'origine (pays de prospection) est indiquée pour les génotypes sauvages. Le numéro inscrit dans la dernière colonne est le code qui sert de référence dans la suite de ce travail.

Certains taxons "sp." (n'ayant pas encore reçu de nom d'espèce) prospectés au Cameroun et au Congo en 1983 (Anthony, 1992), sont également présents. Le matériel étudié est décrit dans le tableau 7.

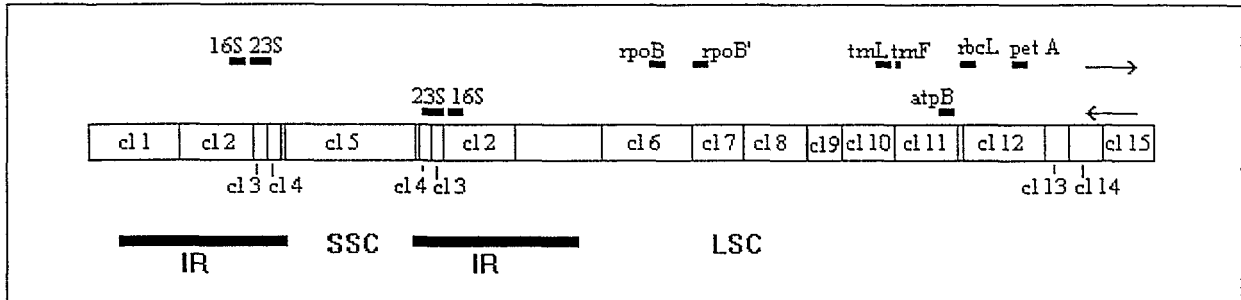
Les travaux présentés dans ce chapitre ont été réalisés à partir de feuilles provenant pour la plus grande partie du centre de Conservation des Ressources Génétiques (CRG) des caféiers africains situé à Man et à Divo en Côte d'Ivoire. Lorsque cela était possible, le prélèvement de feuilles a été réalisé sur les plantes directement issues de prospection ou sur des pieds obtenus par multiplication végétative des génotypes introduits en collection. Pour certains génotypes (*C.liberica* population Koto, *C.sessiliflora*, *C.stenophylla*) l'extraction d'ADN de qualité satisfaisante à partir de feuilles en provenance de Côte d'Ivoire, n'a pas été possible. Dans ces cas précis, nous avons utilisé des échantillons provenant de la "core collection" présente en serre à Montpellier. Il s'agit de plantes issues de graines prélevées en Côte d'Ivoire sur des arbres provenant des prospections. Pour les espèces diploïdes et allogames, les graines ont été collectées au milieu des parcelles afin de limiter le risque d'une éventuelle contamination par un pollen provenant d'une autre population.

Les conditions d'approvisionnement en feuilles provenant de Côte d'Ivoire ont été adaptées de façon à obtenir, après extraction, de l'ADN en quantité suffisante et de bonne qualité. Ainsi, ce sont des feuilles âgées qui sont collectées sur les arbres sélectionnés, les jeunes feuilles s'oxydant plus rapidement. Elles sont simplement placées entre 2 feuilles de papier absorbant puis transportées telles quelles, à température ambiante, par transporteur rapide (4 jours entre le ramassage et la réception au laboratoire). A l'arrivée, elles sont immédiatement triées. Les feuilles oxydées (couleur marron) sont éliminées. Les feuilles en bon état sont lyophilisées durant 72 h, puis broyées en poudre très fine à l'aide d'un broyeur à hélice pour la préparation des farines (CYCLOTECH). Ces poudres sont conservées en pots étanches à -20°C.

#### 3.1.2. Sondes

Un jeu de 15 sondes, correspondant à la digestion de l'ADN chloroplastique de *Lactuca sativa* par l'enzyme de restriction Sac 1 (Jansen & Palmer, 1987), a été utilisé. Ces fragments correspondent à environ 90% du génome chloroplastique de *Lactuca* (150Kb). La carte de restriction de l'ADN chloroplastique de la laitue est représentée dans le schéma ci-dessous (Figure 4). Pour en faciliter la lecture, elle est linéarisée à partir du site de restriction Sac1 situé entre les gènes codant pour *rpl 16* et *rps 3*. Les fragments notés de *cl1* à *cl15* ont des tailles variables, comprises entre 1,8 à 18,8 kb (*cl1* 12,3 ; *cl2* 9,9 ; *cl3* 3,5 ; *cl4* 1,8 ; *cl5* 18,8 ; *cl6*

14,7 ; cl7 7 ; cl8 6,7 ; cl9 3,8 ; cl10 6,9 ; cl11 7,7 ; cl12 10,6 ; cl13 4,6 ; cl14 5,4 et cl15 6,3 kb).



**Figure 4.** Carte de restriction de l'ADNcp de *Lactuca sativa* et localisation des fragments utilisés comme sondes. La zone répétée inversée (**IR**) est repérée par deux lignes noires épaisses sous le schéma. La petite (**SSC**) et la grande (**LSC**) zone en simple copie, sont également indiquées. Quelques gènes chloroplastiques sont localisés (dont les gènes codant pour les ARN ribosomiques). Leur sens de transcription est précisé par les flèches.

### 3.1.3. Choix des enzymes de restriction

Des digestions ont été réalisées par 13 enzymes ayant des sites de reconnaissance à 6 bases (BamHI, BglII, ClaI, DraI, EcoRI, EcoRV, HindIII, KpnI, PstI, SacI, SalI, XbaI, XhoI), sur un échantillonnage d'au moins 4 génotypes (*C.arabica* 325, *C.congensis* 290, *C.pseudozanguebariae* 254, *C.farafaganensis* 329). Un premier choix a été effectué sur la qualité de la digestion produite par ces enzymes. Des membranes de tris et des hybridations ont été réalisées avec les enzymes sélectionnées. Finalement, seules trois enzymes (DraI, EcoRI, EcoRV) ont été retenues pour la qualité des digestions et des hybridations qu'elles engendrent.

## 3.2. METHODES

### 3.2.1. Préparation des membranes

L'ADN est extrait à l'aide d'un protocole modifié (Annexe 1) à partir de celui utilisant le CTAB (Lashermes *et al.*, 1993). Il permet, en règle générale, d'obtenir de l'ADN d'une pureté satisfaisante et de haut poids moléculaire. A partir de 1,5 à 2 g de "poudre" on obtient de 50 à

300 µg d'ADN (la concentration est mesurée, à différentes dilutions, sur un gel d'agarose par rapport à une échelle de concentration connue d'ADN de bactériophage λ).

Quelle que soit l'enzyme utilisée, 7 µg d'ADN sont digérés dans un volume d'au moins 50 µl, en présence du tampon adéquat, et durant 7 à 10 h en étuve à 37°C (1,5U/µg d'ADN, enzymes Boehringer Mannheim). L'ADN digéré est ensuite précipité (éthanol : NaCl 5M, 3 : 0,1), lavé (éthanol 70%), séché (speed-vac) et repris dans 20 µl de TE. Avant dépôt sur gel, une solution (de dépôt) contenant 2 fragments d'ADN de bactériophage λ de 24,8 et 1,5 kb (annexe 2) est ajoutée au TE contenant l'ADN digéré. L'utilisation de ces bandes de poids moléculaire connu, a pour but de faciliter la mise en évidence après hybridation (avec de l'ADN de bactériophage λ) d'éventuelles distorsions de migrations.

Les gels utilisés sont des gels de 20x25 cm (300ml) d'agarose à 1% dans du TBE. Nous utilisons des peignes à 36 dents, qui permettent de déposer 34 échantillons. Les deux pistes situées aux extrémités contiennent des échelles de poids moléculaire (ADN de bactériophage λ digéré par EcoRI et HindIII). La migration est effectuée à 2V/cm durant 16 heures. Après migration le gel est coloré dans un bain de bromure d'éthidium (Bet) (10µg/ml), rincé dans de l'eau permutée, et photographié sur une table U.V. Le transfert est réalisé durant 6 à 8 heures sur des membranes Hybond N+, selon le protocole préconisé par Amersham (dépurination dans HCl 0,25 N durant 20 minutes, rinçage, transfert par capillarité dans du NaOH 0,4N). Les membranes sont ensuite équilibrées dans du 2xSSC. Avant leur utilisation en hybridation (ou entre 2 utilisations), elles sont conservées humides, à 4°C, entre 2 feuilles de papier Wathman dans un sachet plastique scellé.

#### 3.2.2. Préparation des sondes et hybridations

Les sondes chloroplastiques, clonées dans le plasmide PUC 12, ont été fournies par R.K. Jansen sous la forme de précipité d'ADN en présence d'éthanol. Les plasmides contenant les sondes "Cl" sont utilisés pour la transformation de bactéries DH5α (Sambrook *et al.*, 1989). L'ADN plasmidique est extrait puis digéré par l'enzyme de restriction SacI. L'insert est isolé sur gel d'agarose 1% et purifié grâce à des cycles de congélation/centrifugation (annexe 3).

Pour l'ensemble des hybridations avec des sondes chloroplastiques, nous avons utilisé un marquage au dCT<sup>32</sup>P. Nous avons utilisé entre 15 et 30 ng d'insert marqué au <sup>32</sup>P par la technique du "random priming" (kit Mégaprime d'Amersham). Les différentes étapes de préhybridation, hybridation et lavages de stringence sont réalisées en tube dans un four à

hybridation à 65°C (annexe 4). Le temps d'exposition varie de 4 à 72 heures en fonction de l'intensité du signal. Après exposition, les membranes sont deshybridées par traitement dans du SDS 1% bouillant jusqu'à déshybridation complète (signal inférieur à 1 coup par seconde). Chaque membrane peut être réutilisée pour 10 à 15 hybridations successives.

### 3.2.3. Codage des données

Dans un premier temps, ce sont les fragments de restriction qui ont été codés et leur taille estimée. Par la suite, pour plus de rigueur dans l'analyse du polymorphisme, une carte génétique sommaire (mais positionnant l'ensemble des sites polymorphes que l'on a pu mettre en évidence) a été établie. Un codage pour l'ensemble des sites polymorphes a ainsi été effectué. Par rapport à un codage des fragments polymorphes, le codage de sites cartographiés permet d'éviter le biais possible lors d'une analyse en polymorphisme de longueur : 2 "caractères bandes" codés au lieu d'1 "caractère site" réel et la redondance d'un site codé 2 fois lorsqu'il est repérable par 2 sondes contiguës.

Ces deux types de codages (fragments et sites) se complètent, l'un permettant le calcul des poids moléculaires, l'autre étant plus adapté à l'estimation de la divergence.

### 3.2.4. Estimation de la divergence

La divergence nucléotidique a été estimée entre les génotypes deux à deux, à partir du nombre de sites polymorphes ( $N_p$ ) et du nombre total de sites mis en évidence ( $N$ ). Dans le cas d'utilisation d'enzyme de restriction ayant des sites de reconnaissance à 6 bases, et en considérant que le polymorphisme de site de restriction est dû à une seule mutation ponctuelle, la divergence ( $D$ ) s'écrit :  $D = 100 N_p / 6N$ .

La proportion de sites informatifs entre les espèces du genre *Coffea* ( $P$ ) est déterminée comme étant :  $P = N_p \text{ total} / N$ .

La proportion de sites informatifs estimée avec cette formule constitue une estimation du polymorphisme intragénérique, et non pas une divergence entre 2 espèces.



### **3.3 RESULTATS ET DISCUSSION**

Toutes les sondes utilisées ont donné un signal intense sur toutes les espèces de *Coffea* et de *Psilanthus* ce qui traduit une homologie de séquence élevée de l'ADNcp des caféiers avec l'ADNcp de la laitue. Ce résultat est assez général chez les angiospermes et particulièrement chez les *Rubiaceae*, pour lesquels l'ADNcp est suffisamment conservé pour obtenir un signal à l'aide de sondes provenant d'ADN chloroplastique de *Lactuca* (Bremer & Jansen, 1991).

Les 45 couples enzymes/sondes permettent de détecter 169 fragments différents chez les 52 génotypes de caféiers étudiés. Aucun remaniement structurel (large inversion) n'a été repéré entre les taxons. De même, aucune insertion ou délétion portant sur de longs fragments (de l'ordre de 1000pb) n'a été détectée.

Cependant, au moins 4 couples enzyme/sonde mettent en évidence des mutations de type "insertion/délétion" courtes. Par exemple, la sonde Cl4 qui couvre l'extrémité de la région inversée répétée (voir figure 1), révèle un polymorphisme de ce type. Utilisée sur des membranes portant des ADN digérés par EcoRI, la sonde Cl4 s'hybride à une bande située à environ 1200pb et qui possède un poids moléculaire variable. Ce type d'insertion/délétion courte est souvent présent dans des zones de points chauds où les mutations sont "superposées" (Barriel, 1994). La méthode utilisée (gel d'agarose 1%) n'a pas permis la détermination des tailles de ces "indels".

Malgré la présence d'indels courts, l'absence d'insertion ou de délétion portant sur de longs fragments est donc le signe que l'ADNcp des caféiers possède une taille assez conservée.

#### **3.3.1. Estimation de la taille de l'ADNcp des caféiers**

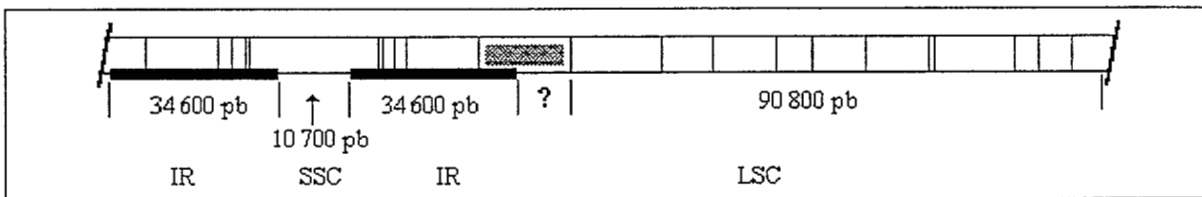
La somme de l'ensemble des fragments différents obtenus pour 1 génotype après digestion avec une seule enzyme et hybridation à un jeu de sondes correspondant à la totalité de l'ADNcp, permet théoriquement de calculer la taille totale de l'ADN chloroplastique.

Cette approche a été appliquée au génotype IF 200 (*C.canephora* 265). La digestion de l'ADN total par l'enzyme EcoRI, suivie de l'hybridation avec le jeu de 15 sondes, a permis de révéler 40 fragments différents. L'addition des poids moléculaires calculés pour chacun des 40 fragments donne pour cet échantillon une taille estimée d'environ 170 kb. Une autre estimation

réalisée pour le même génotype, mais basée sur la somme des 24 fragments différents repérés après digestion avec l'enzyme EcoRV, donne une valeur de 162 kb.

Cette approche a également été appliquée à un autre génotype : l'hybridation de l'ADN de *C.pseudozanguebariae* (254) digéré par EcoRI avec les 15 sondes révèle 39 fragments différents. Leur somme donne une valeur de 168 kb.

La variation entre les différentes estimations est à relier à l'imprécision dans la détermination des poids moléculaires des bandes. Sur l'ADNcp de caféier, les hybridations réalisées à partir d'échantillons digérés par EcoRI produisent les fragments les plus nombreux. Ces fragments sont en moyenne plus petits, ce qui augmente la précision dans la détermination de leur taille. L'estimation réalisée à partir des ADN digérés par EcoRI est, pour cette raison, sans doute plus proche de la réalité. Sur la figure 5 sont portées les tailles estimées pour les différentes zones de cet ADNcp.



**Figure 5.** Représentation schématique des tailles des régions de l'ADNcp de l'IF 200 (*C.canephora*). Les valeurs indiquées sont estimées à partir des résultats de l'hybridation successive des 15 sondes après digestion par l'enzyme EcoRI.

Pour la détermination de la taille de l'ADNcp total, un deuxième facteur doit être pris en compte. En effet, l'ensemble du jeu de sondes utilisé ne correspond qu'à 90% du génome chloroplastique de la Laitue (voir figure 5). Dans la grande région en simple copie, il existe une zone de 15,4 kb, non "recouverte" par le jeu de sondes (Jansen & Palmer, 1987). Cette zone est repérée sur la figure 5 par un "?". La prise en compte du poids moléculaire de cette zone permet de proposer une taille approximative de 180 kb pour le génome chloroplastique des caféiers.

Les caféiers ont un ADNcp de taille relativement élevée, comparée aux tailles déterminées pour d'autres angiospermes. Par exemple, la taille de l'ADNcp de 2 espèces distantes de la famille des *Asteraceae* a été calculée par Jansen et Palmer. Ils donnent une taille approximative de 151 kb pour *Lactuca sativa* et *Bardenasia carophylla* (Jansen & Palmer, 1987). Cette taille peut être décomposée en deux régions inversées répétées d'environ 25 kb, une petite zone en simple copie d'environ 15 kb et une grande région simple copie de 86 kb (Jansen & Palmer, 1987). Il

existe chez les *Asteraceae* une inversion de 22 kb située dans la grande région simple copie, toutefois sans influence sur la taille de l'ADNcp. Chez les *Rubiaceae*, seule la taille de l'ADNcp de *Psychotria bacteriophila* (sous famille des Rubioideae), a été déterminée (Bremer & Jansen, 1991). Selon l'enzyme utilisée pour la digestion de l'ADN, l'estimation varie de 146,6 à 156,8kb (seuls les fragments de plus de 400pb ont été identifiés et comptabilisés). Les IR sont estimés à 26kb, la petite région simple copie entre 19 et 22 kb et la grande région en simple copie à environ 84kb. Chez les caféiers, l'ADNcp semble plus grand que chez *Psychotria bacteriophila*. Cette taille plus élevée est essentiellement liée à l'augmentation de la taille des IR. Dans de nombreux autres cas, les IR semblent intervenir dans une variation de la taille de l'ADNcp. Chez certaines légumineuses il n'existe qu'une copie d'IR et la taille de l'ADNcp varie entre 120 et 140kb, alors qu'au contraire chez le géranium ce sont 2 régions IR de 76kb qui sont à l'origine d'une taille de l'ADNcp de 217kb (Palmer, 1987).

### 3.3.2. Mise en évidence du polymorphisme de l'ADNcp entre les espèces de caféiers

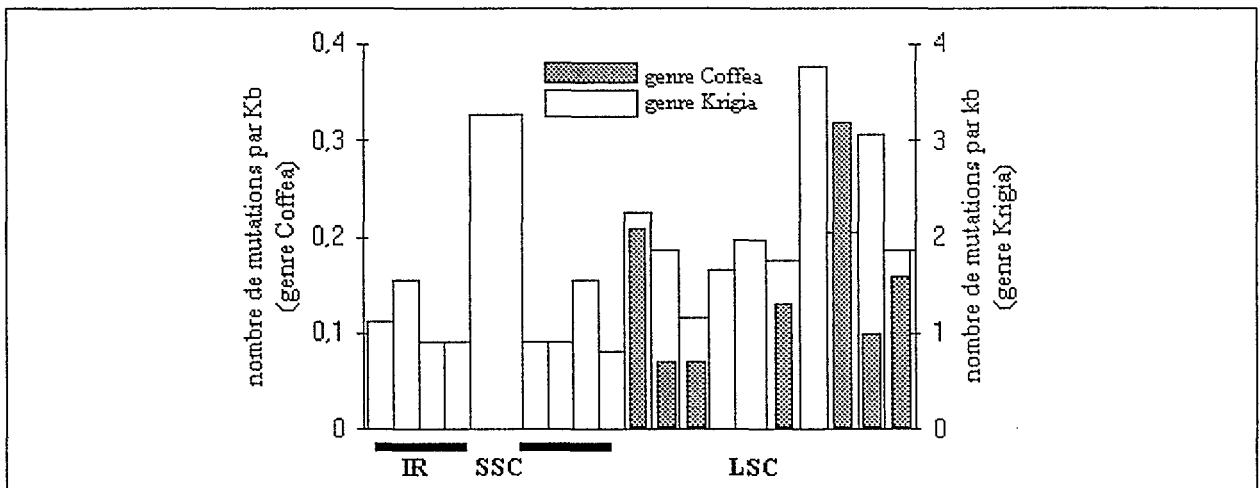
Chez les caféiers, l'ensemble des 45 couples enzyme/sonde révèle 169 fragments qui correspondent à 112 sites de restriction (répartis sur l'ensemble du génome chloroplastique). Entre les 52 génotypes étudiés, 8 des 112 sites repérés sont polymorphes. Les 8 sites révélant du polymorphisme (présence ou absence du site selon le génotype étudié) sont codés dans le tableau 8.

site	enzyme	sonde
A	EcoRI	Cl 6
B	EcoRI	Cl 6
C	EcoRI	Cl 7/Cl 8
D	EcoRI	Cl 11
E	EcoRI	Cl 13/Cl 14
F	EcoRI	Cl 15
G	EcoRV	Cl 6
H	DraI	Cl 13

**Tableau 8** Couples enzyme/sonde(s) mettant en évidence 8 sites polymorphes entre 52 génotypes de caféiers.

La répartition des sites entre les trois enzymes n'est pas homogène : de façon significative la digestion à l'aide de l'enzyme Eco RI révèle le plus de polymorphisme, puisque 6 des 8 sites polymorphes sont révélés par cette enzyme. De même, la répartition des sites polymorphes n'est pas homogène sur l'ensemble de l'ADNcp. Tous les sites sont mis en évidence à l'aide de

seulement 7 sondes situées dans la grande région simple copie bien qu'elle ne représente que la moitié (en taille) de l'ADNcp (figure 6). Certains d'entre eux sont mis en évidence à l'aide de plusieurs sondes contiguës, nous permettant ainsi de les localiser (site E révélé avec EcoRI et Cl13/Cl14 et site C révélé avec EcoRI et Cl7/Cl8).



**Figure 6** . Localisation des 8 sites de restriction polymorphes mis en évidence dans le genre *Coffea*. Pour comparaison, les résultats publiés pour le genre *Krigia* (résultats obtenus avec le même jeu de sondes), sont également indiqués (Jansen *et al.*, 1992). En abscisse sont représentées les différentes zones de l'ADNcp.

### 3.3.3. Hétérogénéité intragénotype de l'ADNcp

*C.millotii* (308) *C.farafanganensis* (329) *C.pervilleana* (238) *C.humblotiana* (330) et *C.bertrandii* (408) possèdent à la fois la bande correspondant à l'absence d'un site, et celles correspondant à sa présence. Il apparaît nettement que lorsque ces deux types de profils coexistent chez un génotype, l'un des 2 profils (correspondant à la présence ou à l'absence du site) est toujours plus intense que l'autre. L'intensité relative des 2 profils sur un même génotype est reliée au nombre de copies portant ou non la mutation. Il est donc possible d'exclure une duplication de cette zone comme mécanisme unique. La coexistence de 2 types de profil (avec des intensités différentes) pour un site donné chez un même individu est donc le signe qu'il existe (chez une même plante) 2 populations d'ADNcp dont l'une est majoritaire.

Cette hétérogénéité est en contradiction avec la présence d'un seul type d'ADNcp comme il est couramment admis (Sugiura, 1992). L'absence (ou la perte) d'hétéroplasmicité est généralement considérée comme liée à la combinaison de 2 phénomènes : transmission au

hasard d'un très faible nombre de copies et multiplication au hasard d'une seule copie dans un génotype donné (Birky, 1990). Pourtant, bien que le mécanisme n'en soit pas expliqué, ce type d'hétérogénéité "intra-plante" de l'ADNcp a déjà été observé (Scott *et al.*, 1991), aussi bien chez des angiospermes que chez des gymnospermes. Chez *Medicago sativa*, il est montré que peuvent co-exister à l'intérieur d'une même plante 2 génomes chloroplastiques (dont l'un est majoritaire) différant par au moins un site de restriction (Rose *et al.*, 1986). De même, certains hybrides entre *Pinus banksiana* et *Pinus contorta* possèdent 2 types d'ADNcp (Govindaraju *et al.*, 1988).

Une origine possible de cette hétérogénéité est la présence exceptionnelle d'héritage biparental (Clegg *et al.*, 1990). Elle a été montrée récemment chez plusieurs espèces (Masoud *et al.*, 1990; Mason *et al.*, 1994; Reboud & Zeyl, 1994). Chez les caféiers, Berthou *et al.* (1983) ont observé sur un croisement interspécifique, à partir de digestions de l'ADNcp isolé, que l'hérédité était maternelle. Ces résultats ont été récemment confirmés sur un croisement interspécifique, entre *C.arabica* et *C.canephora* par la digestion de fragments chloroplastiques amplifiés par PCR (Lashermes, communication personnelle). Toutefois, des phénomènes exceptionnels d'hérédité biparentale ne peuvent pas être exclus.

#### 3.3.4. Détermination des chlorotypes des taxons de caféiers

Pour le matériel végétal étudié, la combinaison des 8 caractères polymorphes permet de définir 11 chlorotypes différents, décrits dans le tableau 9. La présence ou l'absence d'un site de restriction est notée respectivement 1 et 0. Deux de ces chlorotypes (Pi et Pj) sont hétérogènes pour les caractères B et G. Cette hétérogénéité est repérée par un codage 0/1 ou 1/0. Le premier caractère (0 ou 1) indique l'état du caractère pour la population d'ADNcp majoritaire, le second (/1 ou /0) pour la population minoritaire.

Chlorotype	caractères codés							
	A	B	C	D	E	F	G	H
Ca	0	0	0	1	0	0	1	0
Cb	1	0	0	1	0	0	1	0
Cc	1	0	1	0	0	0	1	0
Cd	0	0	0	1	1	0	1	0
Ce	0	0	0	1	0	0	1	1
Cf	0	0	1	0	0	0	1	0
Cg	0	0	0	1	0	0	0	0
Ch	0	1	0	1	0	0	0	0
Ci	0	0/1	0	1	0	1	1/0	0
Cj	0	0	0	1	0	1	1/0	0
Ck	0	0	0	1	0	1	1	0

**Tableau 9.** Etats des 8 caractères polymorphes pour les 11 chlorotypes déterminés.

Les 52 individus étudiés ne sont pas tous discriminés puisque ne sont distingués que 11 chlorotypes différents. Chaque chlorotype peut être porté par un seul individu ou partagé par plusieurs individus de même espèce ou d'espèces différentes. Lorsqu'une espèce est représentée par plusieurs échantillons représentant des groupes de diversité différents, tous portent alors le même chlorotype.

La comparaison des profils de digestion de l'ADNcp isolé provenant de 5 espèces de caféiers (*C.arabica*, *C.canephora*, *C.congensis*, *C.eugenioides*, *C.liberica*) a permis de proposer l'existence d'au moins 3 types d'ADNcp (Berthou *et al.*, 1983). L'augmentation des effectifs (pour ces 5 espèces) nous permet de confirmer ce résultat. De plus, l'élargissement de notre étude à de nombreux autres taxons sauvages permet de distinguer 8 chlorotypes supplémentaires. Le tableau 10 indique la répartition des génotypes entre les 11 chlorotypes.

Chlorotype	effectif	génotypes
Ca	14	<i>C.arabica</i> (211,278,282,236,325) <i>C.eugenioides</i> (263,293,294) <i>C.costatifructa</i> (302) <i>C.kapakata</i> (270,271) <i>C.sp.Mouloundou</i> (256) <i>C.perrieri</i> (268) <i>P.mannii</i> (247)
Cb	6	<i>C.liberica</i> (217,257,260,280,303) <i>C.sp.X</i> (243)
Cc	12	<i>C.canephora</i> (264,265,279,284,285,288) <i>C.congensis</i> (261,290,291,298) <i>C.brevipes</i> (267,276)
Cd	3	<i>C.stenophylla</i> (226,232) <i>C.humilis</i> (258)
Ce	2	<i>C.sp.Mayombe</i> (324) <i>C.sp.N'gongoII</i> (259)
Cf	1	<i>C.sp.N'koumbala</i> (275)
Cg	6	<i>C.pseudozanguebariae</i> (254,255) <i>C.salvatrix</i> (250,251) <i>C.sessiliflora</i> (225,313)
Ch	2	<i>C.racemosa</i> (209,253)
Ci	2	<i>C.millotii</i> (308) <i>C.farafanganensis</i> (329)
Cj	3	<i>C.pervilleana</i> (238) <i>C.humblotiana</i> (330) <i>C.bertrandii</i> (408)
Ck	1	<i>Psilanthus ebracteolatus</i> (248)

**Tableau 10** . Répartition des 52 génotypes étudiés entre les 11 chlorotypes déterminés à partir de l'analyse de 112 sites de restriction de l'ADNcp.

Le chlorotype partagé par *C.arabica* et *C.eugenioides* (Ca) est également partagé par 5 autres taxons ayant des origines géographiques et des caractéristiques très diverses. C'est le chlorotype qui regroupe le plus d'espèces hétérogènes. Il est porté par le taxon autogame récemment prospecté au Cameroun, *C.sp.Mouloundou*. Ce taxon est le seul diploïde sauvage connu qui soit, comme *C.arabica* (tétraploïde), autocompatible. Parmi les 7 taxons qui partagent le chlorotype Ca on peut également noter la présence d'un échantillon de *Psilanthus* : *P.mannii*. En addition à ce résultat, le fait que *P.ebracteolatus* possède un chlorotype qui ne diffère de Ca que pour un seul caractère, confirme la très grande proximité des genres *Psilanthus* et *Coffea*.

Les 10 autres chlorotypes sont partagés par des espèces ayant des origines ou des caractéristiques plus homogènes:

-Le chlorotype Cc, partagé par les 6 génotypes de *C.canephora* et les 4 de *C.congensis*, est également partagé par l'espèce originaire du Cameroun : *C.brevipes*. Les deux premières espèces, non différenciées du point de vue de leur ADNcp, peuvent l'être par l'étude du polymorphisme de restriction de leur ADNmt (Berthou *et al.*, 1983). Elles sont très proches

par leur comportement en croisement (Louarn, 1993) et présentent des similarités du point de vue isoenzymatique (Berthou *et al.*, 1980). Ces 3 taxons sont également proches par leur composition biochimique (Anthony *et al.*, 1993).

- Les espèces malgaches (à l'exception de *C.perieri*) se partagent entre 2 chlorotypes (Ci et Cj), dont la seule différence est une hétérogénéité pour le caractère B.

- Les espèces est africaines, à l'exception de *C.costatifructa* (originaire de Tanzanie), possèdent soit le chlorotype Cg soit Ch, qui ne diffèrent que par 1 caractère polymorphe.

- Le taxon C.sp.X, taxon affine de *C.brevipes* variété hétérocalyx, ne possède pas le même chlorotype que *C.brevipes*. Ce taxon d'origine toujours inconnue n'a manifestement aucune affinité avec les représentants de *C.brevipes* introduits en Côte d'Ivoire (Louarn, 1992). Il partage par contre le chlorotype commun à tous les taxons de *C.liberica*.

- Les 2 espèces endémiques de Côte d'Ivoire (*C.stenophylla* et *C.humilis*), bien que morphologiquement très différentes, possèdent le même chlorotype Cd. De même *C.sp.Mayombe* et *C.sp.N'gongoII* qui sont toutes deux endémiques de la zone forestière côtière du Congo (Manes, 1988), possèdent le même chlorotype Ce.

#### 3.3.5. Estimation de la divergence intragénérique de l'ADNcp

L'estimation de la divergence porte sur 112 sites de restriction (répartis sur l'ensemble du génome chloroplastique) soit 672 nucléotides. En admettant pour l'ADNcp des caféiers, une taille de 180kb, ceci représente un peu moins de 0,5% de l'ensemble de l'ADNcp.

Chez les caféiers, la divergence maximale est observée entre les espèces possédant les chlorotypes Cc (*C.canephora*, *C.congensis* et *C.brevipes*) et Ca (*C.racemosa*). Cinq mutations les distinguent (tableau 9). Si le polymorphisme de site de restriction est considéré être dû à une seule mutation ponctuelle, la divergence nucléotidique maximale observée entre *C.canephora* et *C.racemosa* est inférieure à 1%.

Huit sites polymorphes sur 112 sites de restriction étudiés indiquent une proportion de sites informatifs de 7,2%. Dans le même type d'études RFLP sur l'ADNcp entre espèces congénériques ou appartenant à des genres proches, ce chiffre varie de 2,1 à 9,6% (Bremer & Jansen, 1991). Cette hétérogénéité entre les taux est liée d'une part à la distance génétique entre taxons, mais aussi à l'échantillonnage (et au nombre) d'espèces incluses dans l'étude. A plus haut niveau taxonomique (Famille), la proportion de sites informatifs augmente considérablement. Elle est de 35% chez les *Asteraceae* (pour une divergence nucléotidique



calculée de 3,7%) et atteint 60% chez les *Rubiaceae* (Jansen *et al.*, 1990; Bremer & Jansen, 1991).

### **3.4 BILAN**

Du point de vue de sa structure, l'ADNcp des caféiers semble très conservé, comme d'ailleurs les autres génomes chloroplastiques de *Rubiaceae*. Sa taille, et plus particulièrement la taille de la grande région en simple copie, semble plus grande que celle de l'ADNcp de *Psychotria bacteriophila* (*Rubiaceae*).

Bien que faiblement divergent, l'ADNcp des caféiers est suffisamment variable pour permettre de distinguer 11 chlorotypes entre 52 individus de caféiers appartenant à 27 taxons. Toutefois, le faible nombre de caractères polymorphes et informatifs ne permet pas proposer des hypothèses phylogénétiques. Il semble donc indispensable d'accroître le nombre de caractères informatifs.

Plutôt que d'augmenter le nombre d'enzymes dans une étude concernant l'ensemble de l'ADNcp, il nous a paru plus judicieux de nous intéresser spécifiquement à des zones évoluant plus rapidement, et donc d'étendre notre approche aux régions intergéniques de l'ADNcp. D'autre part, une approche à l'aide d'un outil plus performant que l'étude des sites de restriction mérite d'être envisagée.

#### **4. VARIATIONS DE REGIONS NON CODANTES DU GENOME CHLOROPLASTIQUE DES CAFEIERS**

L'étude de l'ensemble du génome chloroplastique par RFLP a permis de calculer une divergence moyenne inférieure à 1% entre les séquences nucléotiques des ADNcp de caféiers. Afin d'obtenir le maximum d'informations phylogénétiques tout en sauvegardant la "qualité" des caractères (faible homoplasie), Ritland et Clegg suggèrent d'étudier des séquences qui divergent de 5 à 15% entre unités taxonomiques étudiées (Olmstead & Palmer, 1994). L'ADNcp possède des zones plus ou moins variables. Afin d'augmenter le pouvoir de résolution (et indirectement la valeur de la phylogénie établie) nous avons choisi d'étudier des zones dont la divergence entre les caféiers serait supérieure à la valeur moyenne. Notre intérêt s'est ainsi porté vers des régions non codantes. Les amorces permettant d'amplifier de telles zones sont établies à partir des séquences codantes flanquantes. Etant donné la très grande conservation des parties codantes, même les amorces mises au point pour l'amplification chez des espèces éloignées peuvent être employées.

## **4.1. MATERIEL**

### **4.1.1. Matériel végétal**

Le matériel végétal étudié appartient au matériel présenté précédemment. Les échantillons sont décrits dans le tableau 11. On doit se référer au tableau 7 (chapitre 3.1.1) pour les compléments d'informations concernant l'origine et le nom des génotypes étudiés.

Au maximum quatre échantillons par espèce ont été retenus pour les espèces où plusieurs groupes de diversité ont été déterminés. Les problèmes techniques rencontrés au cours des mises au point n'ont pas permis d'inclure à cette analyse 2 espèces de *Coffea* (2 taxons malgaches : *C.pervilleana* et *C.farafanganensis*).

Deux échantillons de *Rubiaceae* appartenant à des genres voisins des caféiers dans la phylogénie proposée par Bremer (Bremer & Jansen, 1991) ont également été introduits dans l'étude. Il s'agit de *Pentas lanceolata* et d'un échantillon de *Gardenia grandiflora*.

Espèce	n° de code
<i>C.arabica</i>	211/236/325
<i>C.bertrandii</i>	408
<i>C.brevipes</i>	267/276
<i>C.canephora</i>	264/265/284
<i>C.congensis</i>	261/290/291
<i>C.costatifructa</i>	302
<i>C.eugenioides</i>	263/294
<i>C.humblotiana</i>	330
<i>C.humilis</i>	258
<i>C.kapakata</i>	270
<i>C.liberica</i>	217/257/260/280
<i>C.millotii</i>	308
<i>C.perrieri</i>	268
<i>C.pseudozanguebariae</i>	254/255
<i>C.racemosa</i>	253
<i>C.salvatrix</i>	250/251
<i>C.sessiliflora</i>	225/313
<i>C.stenophylla</i>	226/232
<i>C.sp.Mayombe</i>	324
<i>C.sp.Mouloundou</i>	256
<i>C.sp.N'gongo2</i>	259
<i>C.sp.N'koumbala</i>	275
<i>C.sp.X</i>	243
<i>P.ebracteolatus</i>	248
<i>P.mannii</i>	247

**Tableau 11.** Génotypes retenus pour l'étude des régions intergéniques de l'ADNcp.

#### 4.1.2. Amorces spécifiques

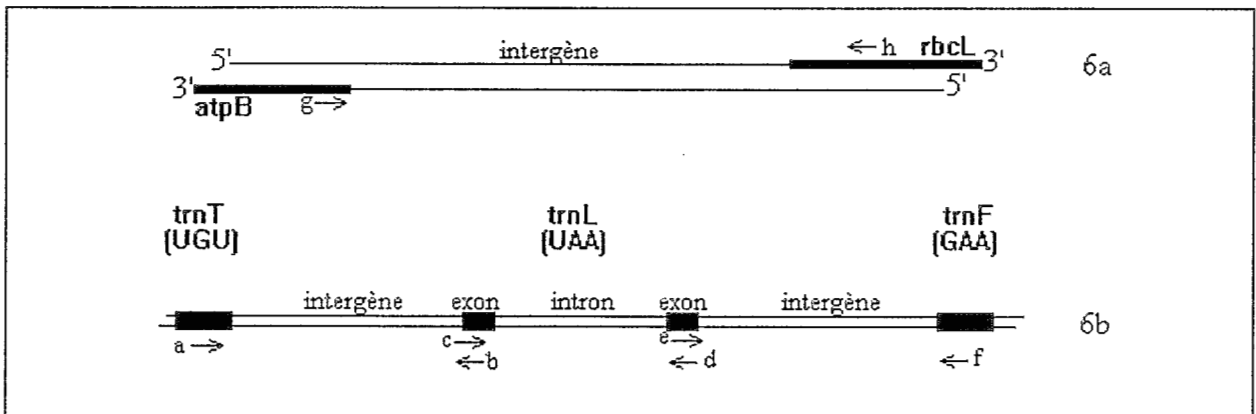
Des amorces permettant l'amplification spécifique de 4 régions non codantes de l'ADNcp ont été testées (Savolainen *et al.*, 1994 ; Taberlet *et al.*, 1991). Elles permettent d'amplifier :

- Les espaceurs intergéniques entre les gènes codant pour les ARN de transfert (*trn*) de la thréonine (T) et de la leucine (L), de la leucine et de la phénylalanine (F)
- L'espaceur intergénique situé entre le gène codant pour la sous unité  $\beta$  de l'ATP synthétase (*atpB*) et le gène codant pour la grande sous unité de la ribulose 1,5-biphosphate carboxylase/oxygénase (*rbcL*)
- L'intron du gène codant pour l'ARN de transfert de la thréonine (*trnT*).

Les amorces utilisées possèdent les séquences suivantes:

- a 5' CATTACAAATGCGATGCTCT
- b 5' TCTACCGATTTCGCCATATC
- c 5' CGAAATCGGTAGACGCTACG
- d 5' GGGGATAGAGGGACTTGAAC
- e 5' GGTTC AAGTCCCTCTATCCC
- f 5' ATTTGAACTGGTGACACGAG
- g 5' GAAGTAGTAGGATTGATTCTC
- h 5' TACAGTTGTCCATGTACCAG

Les positions auxquelles ces différentes amorces s'hybrident (et leur direction) sont repérées par des flèches sur les séquences partielles de l'ADNcp représentées sur les figures 6a et 6b. Les zones traduites sont repérées en gras, la nature de la séquence non codante (intergène, intron) est également précisée.



**Figure 6.** Position des amorces pour l'amplification de 4 zones non codantes. 6a, représentation de la zone non codante intergénique *atpB-rbcL* (d'après Savolainen, communication personnelle). 6b, schématisation des 3 zones non codantes (2 intergènes et 1 intron) adjacentes aux gènes codant pour les ARN de transfert de la thréonine, de la leucine et de la phénylalanine (Taberlet *et al.*, 1991).

## **4.2. METHODES**

### **4.2.1. Amplification par PCR**

Les amplifications sont réalisées dans un volume de 25µl en présence de 10ng d'ADN total, selon le protocole décrit dans l'annexe 6. Pour plus d'homogénéité dans les résultats, on réalise un mélange ("mix") contenant les composés nécessaires à la PCR pour l'ensemble des génotypes utilisés (excepté l'ADN), que l'on distribue dans les tubes préalablement numérotés et contenant l'extrait d'ADN. Dans chaque tube, le mélange final (mix + ADN) est recouvert de 3 gouttes de paraffine liquide afin de limiter l'évaporation.

Les conditions de température d'hybridation de l'amorce avec sa cible et les temps d'élongation ont été optimisés pour chaque couple d'amorce jusqu'à obtenir une bande unique et nette. Elles sont décrites dans le tableau suivant.

	<b>couples e-f, c-f,et c-d</b>	<b>couple a-f</b>	<b>couple g-h</b>
hybridation	55°C / 1 minute	50°C / 1 minute	58°C / 1 minute
élongation	72°C / 2 minutes	72°C / 2 minutes 30 s.	72°C / 2 minutes

Après amplification, l'efficacité de la PCR est contrôlée par dépôt d'un dixième de la solution réactionnelle sur un mini gel d'agarose (annexe 4).

### **4.2.2. Digestion de fragments amplifiés par PCR**

La précipitation du fragment d'ADN amplifié est réalisée directement à partir du mélange réactionnel par ajout de 2,5 volumes d'éthanol. Le mélange est centrifugé, séché, puis repris dans du TE à une concentration d'environ 100 ng/µl. La digestion est réalisée durant 4h à 37°C (ou 65°C pour les digestion par TaqI) en présence de 2 µg d'ADN (fragment purifié), de 6 unités d'enzyme et du tampon adéquat selon les instructions du fournisseur (enzymes de restriction et tampon : Boehringer Manneim).

Les fragments ainsi digérés sont déposés sur gel d'agarose Metaphor à 3%, ou sur gel de polyacrylamide à 10% (annexe 5). Dans les deux cas les gels d'électrophorèse sont photographiés après coloration au bromure d'éthidium.

#### 4.2.3. Séquençage de fragments amplifiés par PCR

Le fragment amplifié à séquencer est isolé sur un gel d'agarose préparatif. Il est ensuite purifié à l'aide de cycles de congélation/décongélation, suivis d'une précipitation de l'ADN contenu dans le mélange (annexe 3). Le fragment d'ADN double brin ainsi purifié est directement utilisé comme matrice pour la réaction de séquence, sans étape de clonage préalable ou d'amplification simple brin.

La réaction de séquence est réalisée par PCR. Il s'agit d'une amplification directe effectuée à l'aide d'une seule amorce (spécifique du fragment à séquencer), à partir de 100ng de matrice purifiée. Outre les 4 déoxynucléotides, le mélange réactionnel contient un mélange des 4 didéoxynucléotides, chacun couplé à un fluorochrome différent. Le protocole de la réaction est décrit dans l'annexe 7. Le produit de cette réaction est déposé sur un gel de polyacrylamide à 6% (annexe 5). La migration s'effectue dans un séquenceur automatique Applied Biosystems qui stocke les informations tout au long de la migration et ceci pour l'ensemble des pistes. Les séquences sont obtenues sous forme de profils quadrichromes.

A partir de chaque fragment "matrice", 2 réactions sont effectuées. Le fragment est séquencé dans les 2 sens à partir de chacune de ses extrémités. Lorsque les fragments à séquencer sont courts (jusqu'à 600pb), cette approche permet leur séquençage complet avec une bonne fiabilité et sans étape intermédiaire de sous-clonage.

#### 4.2.4. Analyse des données

L'alignement multiple des séquences a été réalisé à l'aide du logiciel "CLUSTAL" (Higgins *et al.*, 1992). Des alignements "à la main" ont également été effectués lorsque les séquences étaient très peu divergentes.

A partir des matrices de multi-alignement, nous avons, dans un premier temps, calculé le nombre d'événements mutationnels, leur type et le taux de transition/transversion.

Nous avons ensuite estimé, pour chaque paire de génotypes, la divergence nucléotidique en employant la formule de Kimura à 2 paramètres (Kimura, 1980):

$$d = -1/2 \ln [ (1-2P-Q) (1-2Q)^{1/2} ]$$

P fréquence de transitions entre les 2 séquences

Q fréquence de transversions entre les 2 séquences

La valeur de la divergence calculée à l'aide de cette formule est exprimée en pourcentage. Les "indels" ne sont pas pris en compte. Les calculs sont réalisés à l'aide du programme "DNAdist" du logiciel "PHYLIP" à partir des séquences alignées. Une option permet de préciser le taux de transition/transversion.

La matrice de Kimura permet de construire un dendrogramme par agrégation de type UPGMA<sup>1</sup> (hiérarchie combinatoire de lien moyen). Cette méthode n'est pas "juste" du point de vue phylogénétique lorsque le taux d'évolution est inégal selon les branches, puisqu'elle a tendance à regrouper ensemble tous les individus ayant peu divergé par rapport à leur ancêtre commun. Dans notre cas, elle permet simplement une représentation plus lisible de la matrice de Kimura. Les agrégations ont été réalisées à partir de la matrice de Kimura, à l'aide du programme "neighbor" du logiciel "PHYLIP". C'est l'option "UPGMA" qui a été retenue comme critère d'agrégation.

### **4.3. RESULTATS ET DISCUSSION**

#### **4.3.1 Zones adjacentes aux gènes des ARNt**

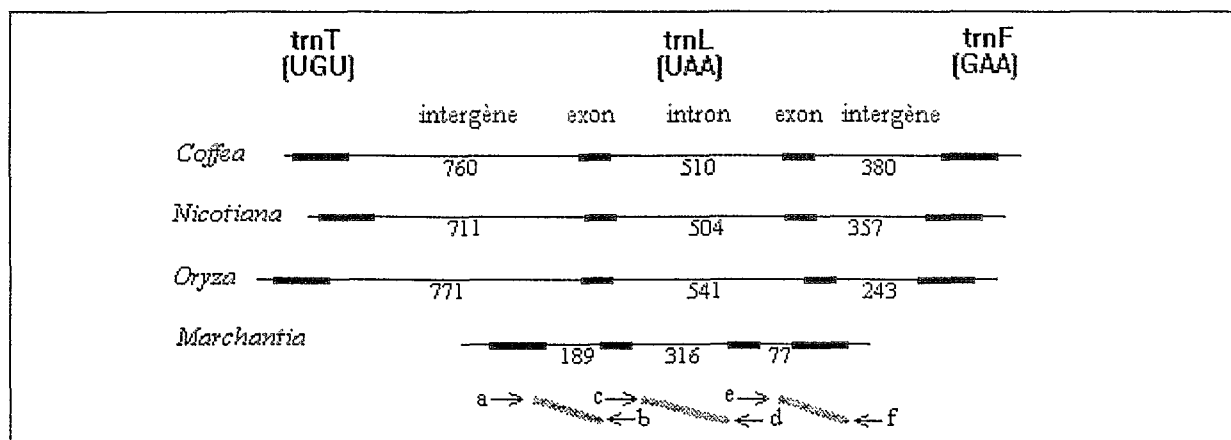
##### taille des fragments

Les six couples d'amorces possibles ont été testés sur plusieurs échantillons de caféiers. Tous les couples produisent une bande unique et spécifique (photo A). Le calcul des poids moléculaires des différents fragments donne une taille d'environ 760 pb pour l'intergène *trnT-trnL*, 380 pb pour l'intergène *trnL-trnF* et 510 pb pour l'intron du *trnL* (toutes les estimations sont données avec la taille des amorces déduites). Ces résultats sont comparés à ceux obtenus par ailleurs pour d'autres espèces (figure 7). Malgré de nombreux essais faisant varier les conditions d'amplification, les résultats obtenus avec l'amorce "a" ont, dans tous les cas, donné des rendements très faibles par rapport aux autres couples d'amorces. Ceci peut être dû à un mauvais appariement de l'oligonucléotide lié à une homologie de séquence insuffisante. Cette amorce n'a donc pas été retenue par la suite.

---

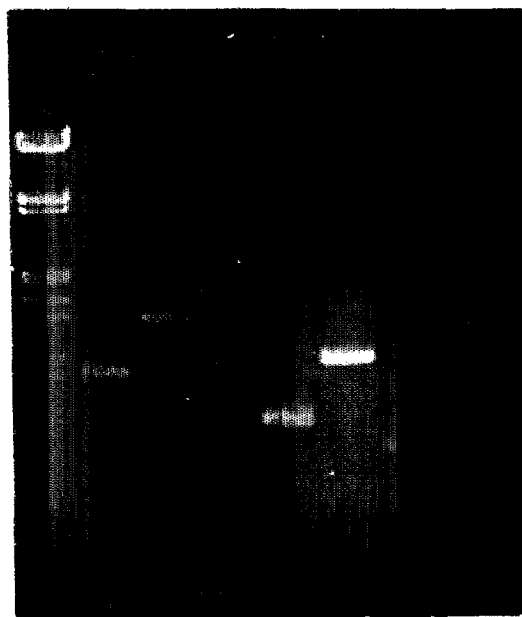
<sup>1</sup> Unweighted Pair-Group Method of arithmetic Average





**Figure 7.** Représentation du polymorphisme de taille observé entre plusieurs genres pour les 3 régions non codantes adjacentes aux gènes des ARN de transfert (le nombre sous ces zones indique la longueur en pb). Les tailles données pour un échantillon de *C.arabica* (n°211) sont des valeurs approximatives déterminées après électrophorèse par comparaison par rapport à un standard de poids moléculaire. Les valeurs indiquées pour les 3 autres genres sont issues des séquences complètes de l'ADNcp du riz (Hiratsuka *et al.*, 1989), du tabac (Shinozaki *et al.*, 1986) et de l'hépatique à thalle *Marchantia polymorpha* (Ohyama *et al.*, 1986).

Bien que les tailles de ces 3 fragments semblent assez variables au niveau intergénérique, aucun polymorphisme intragénérique n'est détecté en gel d'agarose chez les caféiers. La recherche de polymorphisme a donc été entreprise dans un premier temps à l'aide d'enzymes de restriction.

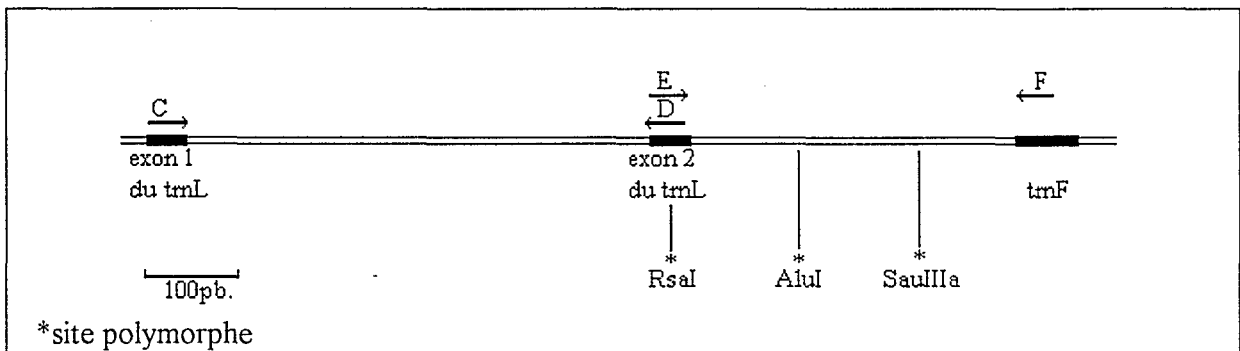


**Photo A** : Profils d'amplification obtenus sur un génotype de *C. canephora* pour les différents couples d'amorces dites "universelles" des zones adjacentes des *trnT*, *trnL* et *trnF*. Puit 1 : marqueur de poids moléculaire ( $\lambda$  digéré par EcoRI et HindIII), 2 : fragment amplifié par le couple (a-b), 3 : (a-d), 4 : (a-f), 5 : (c-d), 6 : (c-f), 7 : (e-f).

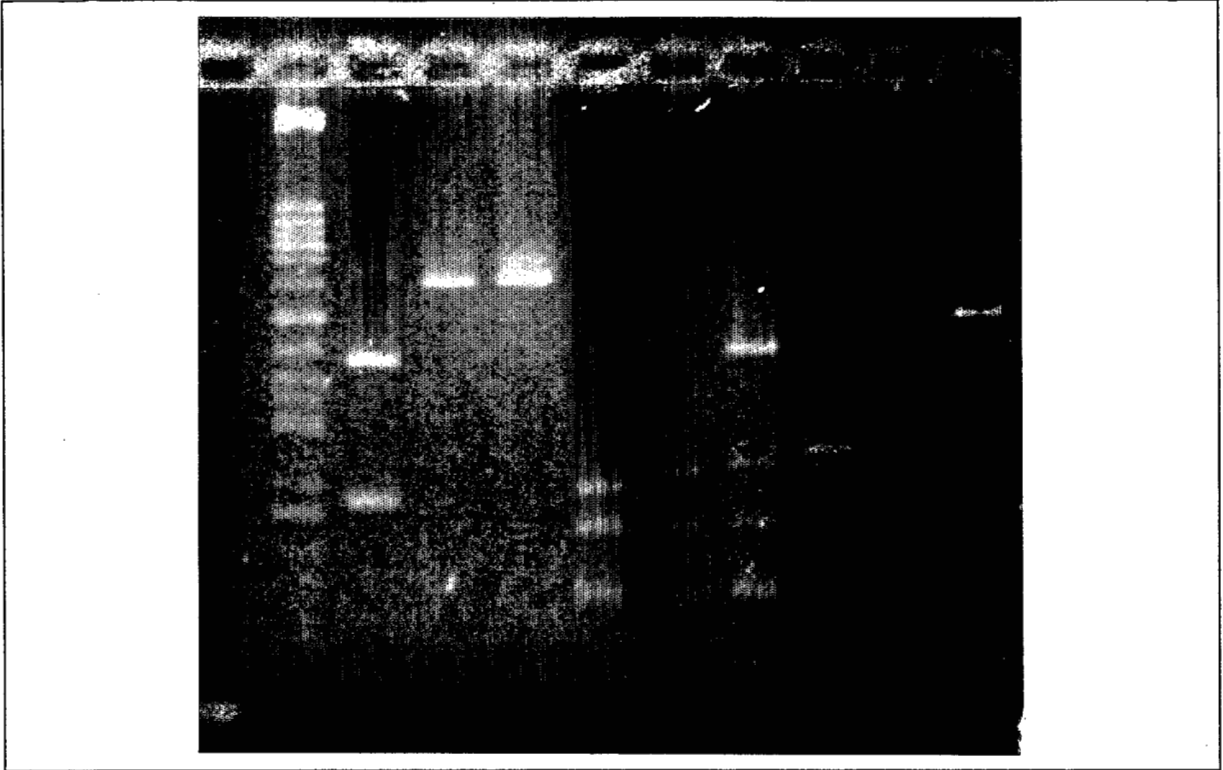
mise en évidence de polymorphisme

L'intron (c-d) et la région intergénique entre *trnL* et *trnF* (e-f) ont été étudiés. Ces fragments après amplification ont été digérés par 6 enzymes de restriction ayant des sites de reconnaissance à 4 paires de bases (HaeIII, AluI, HinfI, RsaI, TaqI et Sau3a). Trois de ces enzymes (RsaI, Sau3a et AluI) mettent en évidence des sites polymorphes entre les génotypes de caféiers. Une série de simples et doubles digestions (voir photos B et C) nous a permis de cartographier ces sites (figure 8).

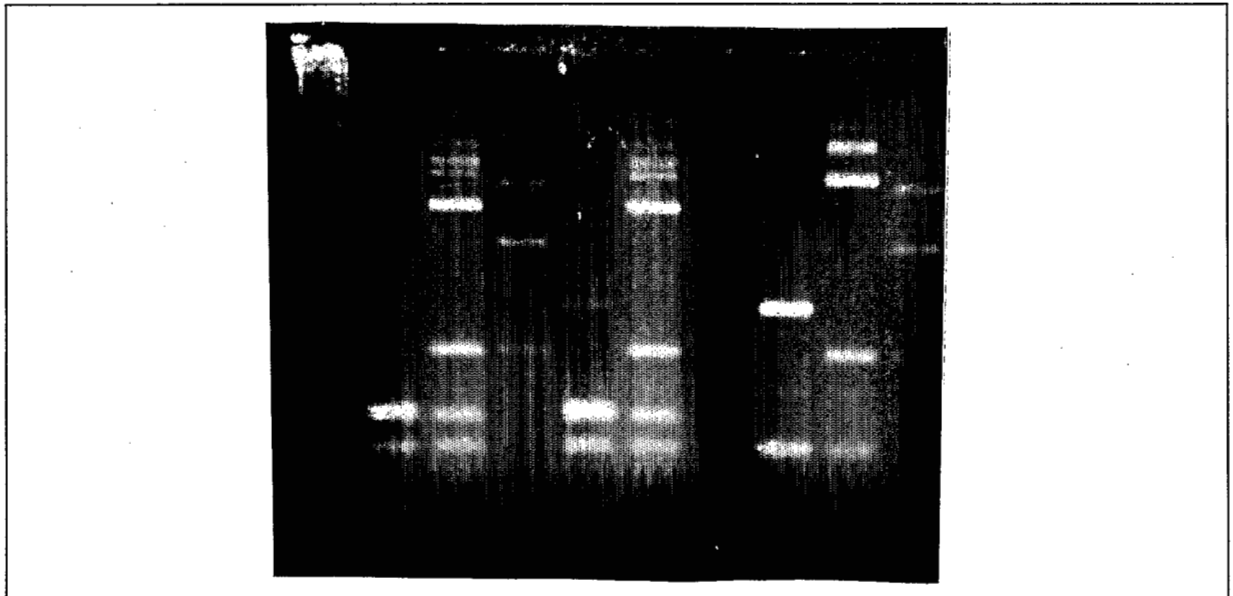
Au moins 2 sites sont localisés sur le fragment intergénique *trnL-trnF*, le site polymorphe RsaI est localisé dans la zone de l'exon 3' du *trnL*. Aucun site polymorphe n'est repéré sur l'intron. la zone de l'intergène entre *trnL* et *trnF* est donc probablement la séquence la plus divergente entre les caféiers. C'est donc cette zone qui a été étudiée par séquençage.



**Figure 8.** Localisation de 3 sites de restriction mettant en évidence du polymorphisme sur le fragment c-f entre les caféiers.



**Photo B** : Localisation des 2 sites polymorphes, AluI et Sau3a. Restriction par AluI (1, 2, 3), par Sau3a (7, 8, 9) et doubles digestions par ces deux enzymes (4, 5, 6) du fragment inter génique *trnL-trnF*, de trois génotypes: *C. congensis* 290 (1, 4, 7), *C. pseudozanguebariae* 254 (2, 5, 8) et *C. arabica* 211 (3, 6, 9).



**Photo C** : Localisation du site polymorphe RsaI. Doubles digestions par AluI et RsaI de l'intergène *trnL-trnF* (1 et 4), de l'intron du *trnL* (3 et 6) et du fragment intron+intergène (2 et 5) pour deux génotypes: *C. congensis* 290 (1, 2, 3) et *C. arabica* 211 (4, 5, 6).

#### 4.3.2. Séquence de l'intergène *trnL-trnF*

Les séquences complètes ont été obtenues pour 40 individus, 38 appartenant au sous genre *Coffea* et 2 au genre *Psilanthus*. Les fragments amplifiés à partir des ADN de *Pentas lanceolata* et de *Gardenia grandiflora* (Rubiaceae) ont également été séquencés.

##### Polymorphisme de taille

Chez les caféiers, la taille de ce fragment varie de 349 (*C.eugenioides* 263) à 383 (*C.sessiliflora*) paires de bases. Ces séquences sont en réalité fortement colinéaires et les différences de taille observées sont exclusivement liées à la présence de 5 délétions. Ces délétions ont des tailles variant de 8 à 30 paires de bases. C'est la faible taille de ces délétions qui explique qu'elles n'aient pas été détectées sur le fragment amplifié entier déposé sur agarose.

##### Mise en évidence des mutations ponctuelles

L'alignement de ces séquences est aisément réalisé "à la main" sur 377 paires de bases et fait apparaître que certains individus sont complètement identiques (tableau 12). Chez les espèces du sous genre *Coffea* nous retrouvons seulement 24 types de séquences présentant chacune une ou plusieurs mutations particulières (figure 9). Au total ce sont 28 types de séquences qui sont distingués à partir des 42 génotypes pour lesquels la séquence a été déterminée.

Code séquence	Génotypes partageant la séquence
Sa	<i>C.arabica</i> 210/211/236, <i>C.sp.Mouloundou</i> 256, <i>C.eugenoides</i> 294
Sb	<i>C.sessiliflora</i> 411
Sc	<i>C.sessiliflora</i> 412
Sd	<i>C.eugenoides</i> 263
Se	<i>C.kapakata</i> 270
Sf	<i>C.pseudozanguebariae</i> 254/255, <i>C.salvatrix</i> 250/251, <i>C.racemosa</i> 253
Sg	<i>C.costatifructa</i> 302
Sh	<i>C.canephora</i> 264/265/284, <i>C.congensis</i> 261/290/291, <i>C.brevipes</i> 267/276
Si	<i>C.sp. X</i> 243
Sj	<i>C.sp. Mayombe</i> 324
Sk	<i>C.sp. Nkoumbala</i> 275
Sl	<i>C.sp. NgongoII</i> 259
Sm	<i>C.liberica</i> 217/260
Sn	<i>C.liberica</i> 257/280 (dewrevrei)
So	<i>C.stenophylla</i> 226/232
Sp	<i>C.humilis</i> 258
Sq	<i>C.millotii</i> 308
Sr	<i>C.bertrandii</i> 408
Ss	<i>C.humblotiana</i> 330
St	<i>C.perrieri</i> 268
Su	<i>Psilanthus mannii</i> 247
Sv	<i>Psilanthus ebracteolatus</i> 248
Spen	<i>Pentas lanceolata</i>
Sgar	<i>Gardenia grandiflora</i>

**Tableau 12.** Regroupement des génotypes en fonction de la séquence de leur inter-gène *trnL-trnF*. Les codes (Sa, Sb...) sont ceux utilisés dans l'alignement présenté figure 9.

La comparaison de ces 28 types de séquences met en évidence un total de 41 mutations dont 22 transitions, 12 transversions et 7 délétions courtes (1 à 30 paires de bases). Neuf de ces mutations sont spécifiques de l'échantillon de *Gardenia grandiflora*. Entre les 40 génotypes de caféiers (*Coffea* et *Psilanthus*), ce sont 32 mutations, dont seulement 13 sont phylogénétiquement informatives, qui distinguent les 26 types de séquences. La proportion de transitions/transversions est de 16/10.

Les séquences des deux échantillons appartenant au genre *Psilanthus* (Su et Sv) sont très proches des séquences des *Coffea*. Seul *P. mannii* possède une délétion courte de 8 paires de bases et 4 mutations ponctuelles qui ne sont partagées par aucun autre taxon (Figure 9).

Sa	TTTGATCCCC	CAACTATTTA	TCCTATCCCC	CTTTCGTTAG	CGGTTCAAAA	AACCTTATTC	ATTTACTCTA	TTCTCTTAGA	AATCGATCTG	GACGGAAAAG
Sb	-----	--0-----	-----	-----	-----	-----	-----	-----	-----	-----
Sc	-----	--0-----	-----T-	-----	-----	-----	-----	-----	-----	-----
Sd	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Se	-----	-000000000	0000000000	0-----	-----	-----	-----	-----	-----	-----
Sf	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sg	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sh	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Si	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sj	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sk	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sl	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sm	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sn	-----	-----	-----	-----	--00000000	000-----	-----	-----	-----	-----
So	-----T	-----	-----	-----	-----	-----	-----T-	-----T-	-----	-----
Sp	-----T	-----	-----	-----	-----	-----	-----T-	-----T-	-----	-----
Sq	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sr	XXXXXXXXXX	XXXXXX--	-----	-----	-----	-----	-----	-----	-----	-----
Ss	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
St	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Su	-----	-----	-----	-----	-----	-----	-----	-----	00000000--	-----
Sv	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Spen	XXX-----	-----G	-----	-----	-----	-----	-----	-----	-----	-----
Sgar	XXXXXXXX--	--0-----	-----T	-----	-----	T-----	-----	-----A-	-----	-----
Sa	CCCTTTTCTT	ATCACAAATC	TTGTGTTATT	TATGATATAC	ATATAAATGA	ACATCTTTGA	GCAAGAAATA	CCCATTGAA	TGGTTTACAA	TCGATATAAC
Sb	-----	-----	-----	-----	-----	-----	-----	-----C	-----	-----
Sc	-----	-----	-----	-----	-----	-----	-----	-----C	-----	-----
Sd	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Se	-T--GG--	-----	-----	-----	-----	-----	-----CT-	-----	-----	-----
Sf	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sg	-----	-----	-----	-----	-----	-----	-----	T-----	-----	-----
Sh	-T-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Si	-T-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sj	-T-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sk	-T-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sl	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sm	-T-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sn	-T-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
So	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sq	-----	-----	-----	-----	-----	-----	-----	-----G	-----	-----
Sr	-----	-----	-----	-----	-----	-----	-----	-----G	-----	-----
Ss	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
St	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Su	-----	-----	-----	-----	-----	-----	T-----	-----	-----	-----
Sv	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Spen	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sgar	-----	-----	-----	-----	-----000	0000000--	-----	-----	-----	-----

**Figure 9 .** Séquences nucléotidiques complètes de l'espaceur intergénique situé entre l'extrémité 3' du *trnL* (UAA) et le *trnF* (GAA) chez les caféiers et 2 autres Rubiaceae. Les codes (Sa, Sb....) font référence au tableau 12. Un tiret correspond à un nucléotide identique à la séquence de référence (*C.arabica* 211), 0 à une délétion et X à une position non déterminée.

Sa	TATTCATACT	GAAACTTACA	AAGTACTCTT	TTTTAAGATA	CAAGAAATTC	TAGTACCTAG	ATAAAATTTT	GTAATCCCCT	TTCCTTCTTT	TAATTGACAT
Sb	-----	-----	-----	-----	-----	-----	C-----	-----	-----	-----
Sc	-----	-----	-----	-----	-----	-----	C-----	-----	-----	-----
Sd	-----	-----	-----	-----	-----	-----	-----	-----0	0000000000	0000000000
Se	-----	-----	C-----	C-----	-----	-----	C-----	-----	-----	-----
Sf	-----	-----	-----	-----	-----	-----	C-----	-----	-----	-----
Sg	-----	-----	-----	-----	-----	-----	C-----	-----	-----	-----
Sh	-----	-----	-----	C-----	-----	-----	C-----	-----	-----	-----
Si	-----	-----	C-----	C-----	-----	-----	C-----	-----	-----	-----
Sj	-----	-----	C-----	C-----	-----	G-----	C-----	-----	-----	-----
Sk	-----	-----	-----	C-----	-----	-----	C-----	-----	-----	-----
Sl	-----	-----	C-----	C-----	C-----	G-----	C-----	-----	-----	-----
Sm	-----	-----	-----	C-----	-----	G-----	C-----	-----	-----	-----
Sn	-----	-----	C-----	C-----	-----	-----	C-----	-----	-----	-----
So	-----	C-----	-----	-----	-----	-----	C-----	-----	-----	-----
Sp	-----	C-----	-----	-----	-----	-----	C-----	-----	-----	-----
Sq	-----	-----	-----	-----	-----	-----	C-----	-----	-----	G-----
Sr	-----	-----	-----	-----	-----	-----	C-----	-----	-----	G-----
Ss	-----	-----	-----	-----	-----	-----	C-----	A-----	-----	G-----
St	-----	-----	-----	-----	-----	-----	C-----	-----	G-----C	-----
Su	C-----	-----	-----	C-----	-----	-----	C-----	-----	GT-----	-----
Sv	-----	-----	-----	C-----	-----	-----	C-----	-----	-----	-----
Spen	-----	-----	-----	C-----	C-----	G-----	XXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
Sgar	C-----	-----	-----	C-----	-----	-----	C-----	-----C	-----	-----
Sa	AG0CCCCCTT	TTCTCATAAA	ATGAGGATGC	TACATTGGGA	CTGGTCGGGA	TAGCTCAGAT	GTAGAGCAG	AGGACTG	-----	-----
Sb	-0-----	-----	-----	-----	-----	-----	-----	-----AAATCC	-----	-----
Sc	-0-----	-----	-----	-----	-----	-----	-----	-----AAATCC	-----	-----
Sd	0000000000	-----	-----	-----	-----	-----	-----	-----	-----	-----
Se	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sf	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sg	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sh	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Si	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sj	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sk	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sl	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sm	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sn	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
So	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sp	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sq	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sr	-0-----	C-----	-----	-----	-----	-----	-----	-----	-----	-----
Ss	-0-----	0-----	-----	-----	-----	-----	-----	-----	-----	-----
St	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Su	-0--T-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Sv	-0-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
Spen	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXX	-----	-----
Sgar	--A-----T--	-----	-----	-----	-----	-----C-	-----	-----AAATCCTCG	-----	-----

Figure 9 (suite) . Séquences nucléotidiques complètes de l'espaceur intergénique situé entre l'extrémité 3' du *trnL* (UAA) et le *trnF* (GAA) chez les caféiers et 2 autres Rubiaceae.



Les séquences de *Gardenia grandiflora* (368pb) et de *Pentas lanceolata* (322pb) ont montré également une très bonne colinéarité avec les séquences des caféiers. La séquence du *Pentas*, pour des raisons techniques, est toutefois très courte. D'autres alignements ont été réalisés à partir de séquences publiées. Bien que l'alignement soit relativement aisé, le nombre de mutations mis en évidence est alors beaucoup plus important. L'alignement de la séquence de *C.arabica* avec *Petunia hybrida*, deux espèces appartenant à des familles différentes, met en évidence 10 indels de 1 à 5pb et 53 mutations ponctuelles de type transition/transversion.

<i>C.arabica</i> 211	CCCTTT0CGTTAGCGGTT0CAAAAAACCTTAT00000TCATTTACTCTATTCTCTTAGAAATCGATCT
<i>Petunia hybrida</i>	-----T-C-----CC-----TT0-0-----CTTTC-----CC-----T-----G--T--
<i>C.ara.</i>	GGACGGAAAAGCCCTTTTCTTATCACAAATCTTGTTATTTATGATATACATATAAATGAACATCTT
<i>P.hyb.</i>	-AG--T---T-G---C-----G-----A---A-----C-----G--0-----0---
<i>C.ara.</i>	TGAGCAAGAAATACCCATTTGAATGGTTTACAATCGATATAACTATTCATACTGAAACTTACAAAGTA
<i>P.hyb.</i>	-----G---C--T---0---C-A--CC-G---A---C-T--C-----C
<i>C.ara.</i>	CTCTTTTAAAGATACAAGAAATCTAGTACCTAGATAAAATTTGTAATCCCCTTTCCTTCTTTTAA
<i>P.hyb.</i>	A-----G-----CG-----GGAT-T-----A---A-TCTA-T---G0-C--T----
<i>C.ara.</i>	TTGACATAG0CCCCCTTTTCTCATAAAATGAGGATGCTACATTGGGACT
<i>P.hyb.</i>	-----A---AG--C---A-----A---G-----A-

**Figure 10.** Séquences nucléotidiques partielles de l'espaceur intergénique situé entre l'extrémité 3' du *trnL* (UAA) et le *trnF* (GAA) chez *Coffea arabica* (Rubiaceae) et *Petunia hybrida* (Solanaceae). Un tiret correspond à un nucléotide identique à la séquence de référence (*C.arabica* 211), 0 à une délétion.

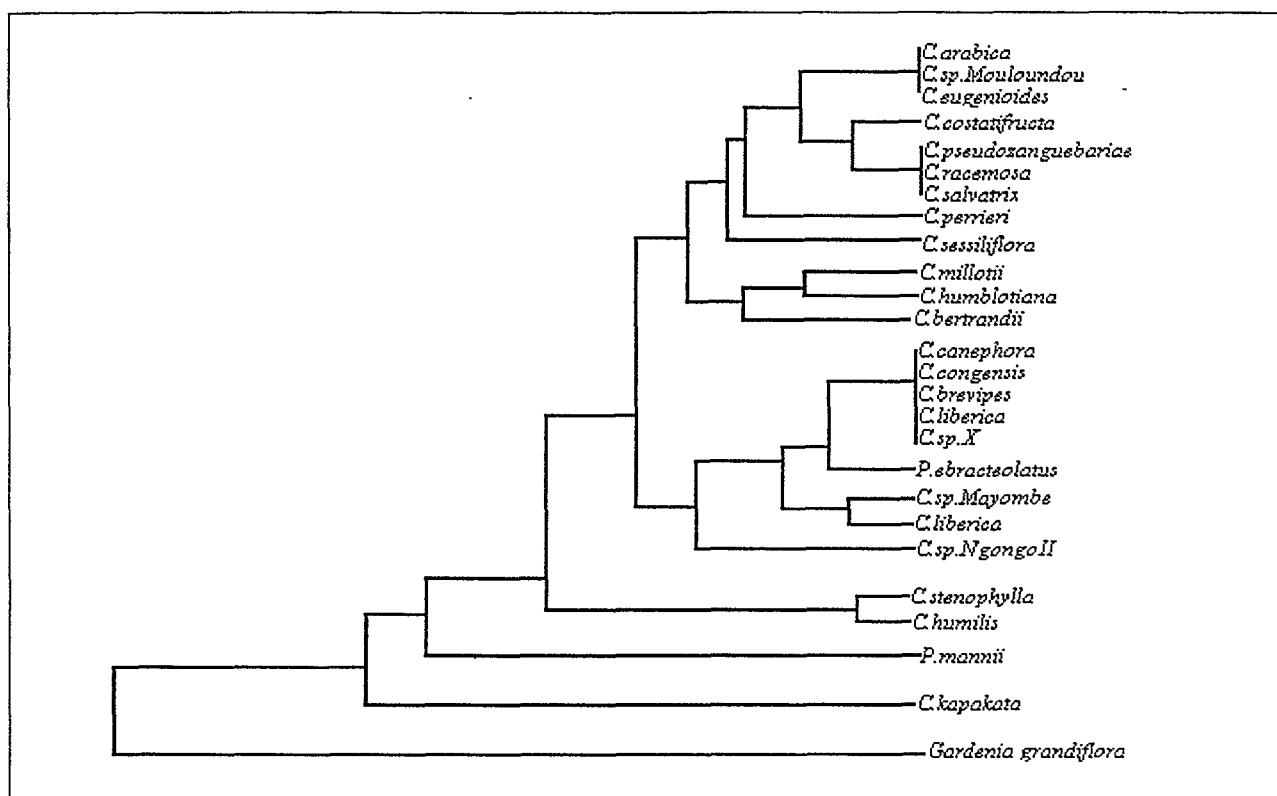
#### Localisation des mutations ponctuelles

Sur l'alignement de 377pb, la totalité des mutations présentes chez les caféiers est localisée entre les positions 9 et 106 d'une part, et 147 et 330 d'autre part. Aucune zone de "point chaud" n'est détectée.

#### Divergence intragénérique

La divergence nucléotidique entre chaque paire de taxons a été calculée à partir de l'alignement sur 377pb avec le programme "DNAdist" de PHYLIP. La distance de Kimura à 2 paramètres a été utilisée, la valeur de 1,8 (22/12) a été donnée pour l'option taux de transition/transversion.

La matrice exprimant la divergence nucléotidique est représentée sous forme de dendrogramme après une agrégation de type UPGMA (figure 11).



**Figure 11.** Dendrogramme construit par analyse UPGMA, à partir de la matrice de Kimura entre 21 types de séquences.

La divergence maximale est de 3,7 %. Elle a été observée entre *Gardenia grandiflora* et *Coffea kapakata*. La position particulière de *C.kapakata* et de *P.mannii* sur le dendrogramme indique une divergence élevée de leurs séquences par rapport aux autres caféiers. Cette divergence est produite par plusieurs mutations propres à chacune de ces séquences. Par contre, l'échantillon de *P.ebracteolatus* qui possède une séquence intergénique très peu divergente par rapport à l'ensemble des caféiers, est placé dans une position très proche de caféiers du sous genre *Coffea*. Chez les *Coffea* au sens strict et excepté *C.kapakata*, la divergence maximale est de 2,0 %. Elle est observée pour le couple *C. stenophylla*, *C. humboldtiana*.

Les résultats obtenus confirment le taux de divergence relativement faible estimé lors des premiers résultats de restriction de ce fragment par des endonucléases. Des travaux sur ce fragment montrent des taux très variables en fonction des genres étudiés (Gielly, 1994). Ces résultats sont comparés dans le tableau 13. Dans le genre *Echium* la divergence maximale

observée entre les séquences de l'intergène est de 1,4% (Böhle *et al.*, 1994). Entre certaines espèces des genres *Fraxinus* et *Alnus*, aucune variation interspécifique n'a pu être détectée. Dans ces genres, la divergence atteint au maximum les valeurs de respectivement 1,14 et 1,52% (Gielly, 1994). Par contre dans le genre *Gentiana*, elle atteint la valeur maximum de 6,4% entre *Gentiana punctata* et *Gentiana alpina* (calculée sur un alignement partiel de 195pb, d'après Gielly, données non publiées).

Le taux de divergence intragénérique observé sur l'intron du gène *trnL* dans le genre *Gentiana* est comparable à celui observé dans le genre *Euphorbia*. Il est respectivement 8,9 fois supérieur à celui observé dans le genre *Fraxinus*, et 5,25 fois supérieur à celui observé dans le genre *Alnus* (Gielly, 1994). Chez les caféiers, le taux de divergence de cette zone n'a pas été calculé. Toutefois au vu des résultats de l'analyse RFLP, on peut estimer qu'il est plus faible que pour la zone intergénique.

Genre	nombre d'espèces	Intron du gène <i>trnL</i>		Espaceur intergénique entre <i>trnL</i> et <i>trnF</i>	
		longueur de la séquence	% maximum de divergence	longueur de la séquence	% maximum de divergence
<i>Alnus</i> <sup>(1)</sup>	4	527	1,15%	198	1,52%
<i>Coffea</i>	22	510	non déterminé	380	3,33% <sup>(3)</sup>
<i>Echium</i> <sup>(2)</sup>	4	393	0,25%	324	1,4%
<i>Fraxinus</i> <sup>(1)</sup>	5	417	0,72%	262	1,14%
<i>Gentiana</i> <sup>(1)</sup>	8	386	5,98%	195 (partielle) <sup>(4)</sup>	6,4% <sup>(4)</sup>
<i>Quercus</i> <sup>(1)</sup>	2	476	2,31%	168	0,6%
<i>Swertia</i> <sup>(1)</sup>	2	380	1,05%	225	1,33%

**Tableau 13.** Divergence intragénérique maximale observée (en %) pour 2 des régions non codantes, adjacentes aux *trnL* et *trnF*. <sup>(1)</sup> d'après (Gielly & Taberlet, 1994), <sup>(2)</sup> d'après (Böhle *et al.*, 1994). <sup>(3)</sup> Divergence observée entre *C.kapakata* et *C.stenophylla*. <sup>(4)</sup> D'après Gielly communication personnelle (données non publiées).

Plusieurs hypothèses peuvent être proposées pour expliquer l'hétérogénéité entre genres des taux de divergence des séquences non codantes de l'ADNcp : tout d'abord l'ancienneté des différents genres, ou plutôt la date de radiation des espèces considérées à l'intérieur du genre. Même lorsque l'horloge moléculaire n'est pas calibrée faute de données paléontologiques, il semble que les phylogénies moléculaires respectent l'ordonnancement dans le temps (Nee *et al.*, 1994). L'ancienneté de la radiation serait donc positivement corrélée avec le taux de divergence des séquences (Gielly, 1994). Une autre explication possible concerne les vitesses d'évolution qui seraient variables selon les lignées considérées. Ce résultat a été montré sur des séquences codantes comme *rbcL* (Doebley *et al.*, 1990), où l'on propose des taux de 10<sup>-10</sup> à 10<sup>-11</sup> substitutions/site/an (Chase *et al.*, 1993), voire beaucoup plus faibles pour certaines

espèces pérennes (Frascaria *et al.*, 1993). Il est probable que cette variabilité des taux d'évolution, en fonction des lignées considérées, existe également sur des séquences non-codantes.

### Divergence intergénérique

La divergence des séquences de l'intergène *trnL-trnF* a également été estimée à un niveau taxonomique plus élevé à partir d'alignements réalisés entre séquences publiées. Entre le genre *Coffea* (Asteridae, famille des *Rubiaceae*) et le genre *Petunia* (Asteridae, famille des *Solanaceae*) la divergence calculée selon l'indice de Kimura à 2 paramètres est de 28,6%. Entre *Alnus incana* et *Quercus rubra* 2 dicotylédones ligneuses, cette divergence est de  $11,6 \pm 3,5\%$  (Gielly & Taberlet, 1994). Entre 2 Monocotylédones *Crimum jagus* et *Scilla hyacinthina* elle est de 14,15% (données calculées à partir de séquences présentes dans la base de données EMBL). A des niveaux taxonomiques supérieurs au genre ou à la famille, cette séquence peut donc se révéler trop polymorphe pour établir des phylogénies. Pour ce type d'analyses, se sont les séquences qui divergent de 5% à 15% au maximum entre unités taxonomiques (Olmstead & Palmer, 1994) qui sont étudiées. Dans la majorité des cas, il s'agit de la séquence codante *rbcL* (Chase *et al.*, 1993). La divergence des séquences de l'intergène *trnL-trnF* permet par contre de résoudre des niveaux phylogénétiques beaucoup plus bas que la séquence de *rbcL* (Gielly & Taberlet, 1994).

### **4.3.3. Fragment *atpB-rbcL***

#### Taille du fragment

Le fragment d'ADN amplifié a une taille d'environ 960 paires de bases (amorces comprises). Cette taille correspond à celle du fragment intergénique *atpB-rbcL* et aux 120 premières paires de bases (partie 5' codante) du gène *rbcL*. Après électrophorèse sur gel d'agarose à 1%, aucune différence de taille de ce fragment entre les différents génotypes étudiés ne peut être déterminée.

Une recherche a été effectuée dans les bases de données de séquences biologiques. Une seule séquence du genre *Coffea* a été trouvée dans la base de donnée EMBL sous la référence "CHCARBCLA" (séquence soumise par Manen et Savolainen, réalisée sur un individu de *C.arabica*). Cette séquence a une taille de 921 paires de bases qui est en accord avec celle que nous avons déterminée sur gel ( $921 \text{ pb} + 2 \text{ amorces de } 20 \text{ et } 21 \text{ pb} = 962 \text{ pb}$ ).

Mise en évidence du polymorphisme

Un polymorphisme de restriction a été recherché à l'aide d'enzymes qui possèdent des sites de reconnaissance à 4 paires de bases. Trois enzymes (RsaI, TaqI et HaeIII) révèlent du polymorphisme. Leurs sites de coupure sont repérés sur la séquence de *C.arabica* (figure 12). Chez l'ensemble des caféiers étudiés, ces enzymes permettent d'obtenir 24 fragments distincts (à partir du fragment amplifié par PCR), dont 15 sont polymorphes.

```

1  TAAAATATGT CAAATTTTTT GCGAAAATTA AATTATCGAA TTCAAAACAA ATATCCAATA
61 GCGCGTCGAT CAGTTAATTC AATTTATTCT TCAATAAGAA TTGGGAATAA GCTCTCCATT
121 TGGTTGGTAC CATTCAATCG AATCCAACCTC AATTGTTTCT TTATTCAAAC AATGAATTTT
181 CAAATTCAT CAATCTATTT TATTTTTAGA ATTTCAAGTG GATGAATAGG AATGTTGAGA
241 AGGTCTTTCA TTTATCTATC ATTATAGATA ATCCCATGTG TATTATCCAT TCTAGGAATT
301 CGAACCTGAA CTCTATTTAC ATTACGATTC ATTATTTCTA TCCCATTGGA TTCTTTTATT
361 CAGCATATCG ATTACGCGAG CCTTTTCATT TCTTTTATC CTATATCGCG AATTCTGTAC
421 ATTTTCACAA TCTAGGAATT TACATATACA ACATATAGCG CTGTCAAGGG AAATTTCTTA
481 TTAGTTAGTT ATTTTGATTC CAAAGGGGTA AAAAAGAAAA TTTGGGTTGC GCTATATATA
541 TATGAAAGAG TATACAATAA TGATGTATTT GGCAAATCGC ATGATCTGGT CAATAATTAA
601 TAATCAAACA TTCTGATTAG TTGATAATAT TAGTATTAGT TGGGAATTTT GTGAAAGATT
661 CTTGTGAAAA GTTTAATTAA CAGCTAATTC GTGTCGAGTA GACCTTGTTG TTGTGAGAAT
721 TCTTAATTCA TGAGTTGTAG GGAGGGGATT ATGTCACCAC AAACAGAGAC TAAAGCAAGT
781 GTTGGATTCA AAGCTGGTGT TAAAGAGTAC AAATTGACTT ATTATACTCC TGAATACGAA
841 ACCAAAGATA CTGATATCTT GGCAGCATTC CGAGTAACTC CTCAACCGGG AGTTCCACCT
901 GAAGAACGGG GGGCCGCGGT A

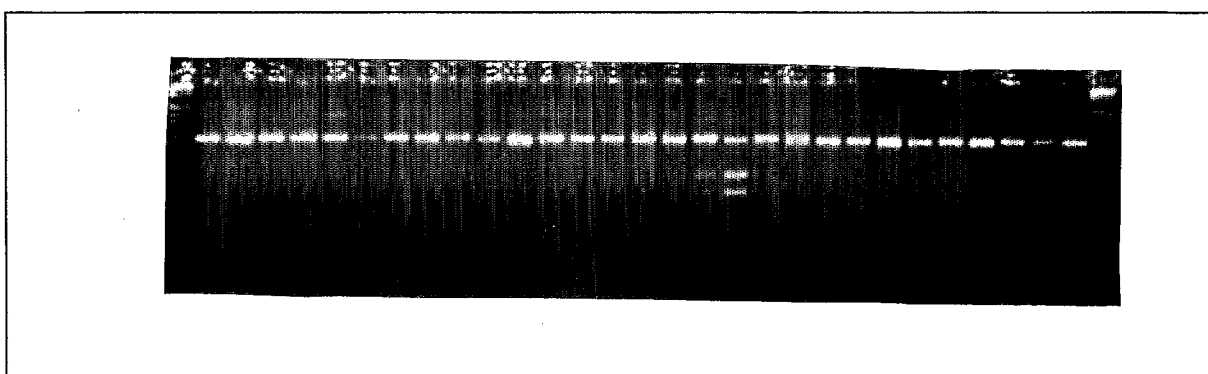
```

**Figure 12.** Séquence de l'intergène *atpB-rbcL* (amorces g-h) de *C.arabica* (ref "CHCARBCLA"). Les sites de restriction de 3 enzymes sont repérés : Rsa I GTAC, Taq I TCGA, Hae III GGCC

Hétérogénéité de l'ADNcp

L'enzyme Hae III permet de mettre en évidence un polymorphisme de site de restriction chez les deux individus de *C.pseudozanguebariae*. Ce polymorphisme s'accompagne d'une hétérogénéité. Nous observons en effet chez ces 2 échantillons, à la fois la bande correspondant à l'absence du site et les 2 bandes correspondant à sa présence (photo D). L'amplification par PCR de ce fragment, suivie de sa digestion, a été renouvelée à trois reprises pour confirmer ce résultat. Le profil observé est toujours le même, il ne s'agit donc pas d'un artefact lié à une erreur de "recopiage" de la Taq polymérase ou à une mauvaise digestion. Il existe donc au moins deux copies différentes de cette zone : soit par duplication d'un fragment d'ADN chloroplastique soit parce qu'il existe deux types d'ADN chloroplastique. La PCR étant une technique très sensible, elle permet l'amplification de cette zone "mutée" même à partir d'un

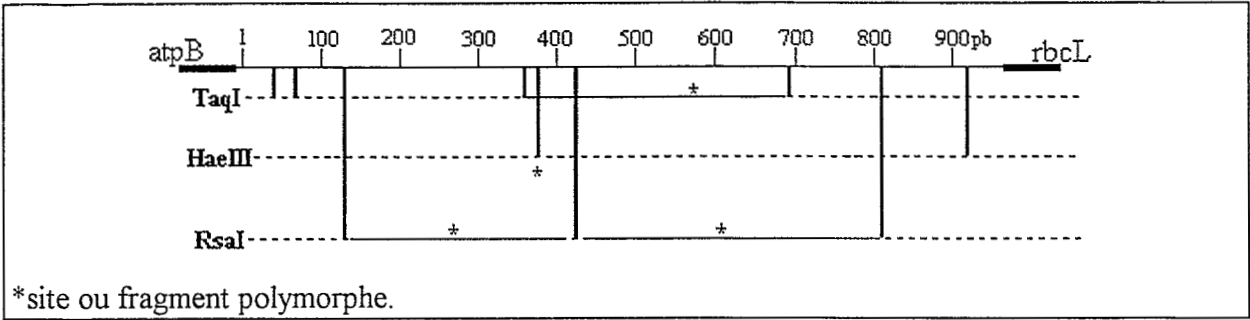
très petit nombre de copies. Ce résultat a également été observé par RFLP sur l'ADNcp total pour 5 taxons (voir paragraphe 4.2.3). La mise en évidence de cette situation à l'aide de 2 techniques différentes (PCR et Southern blot) confirme la présence d'une hétérogénéité de l'ADNcp. Cette hétérogénéité est généralement liée à une hétérogénéité des intensités des 2 profils superposés (présence ou absence du site). L'intensité relative de chacun des 2 profils sur un même génotype signifie probablement qu'il existe toujours une population d'ADNcp majoritaire dans un génotype donné.



**Photo D** : fragments intergéniques *atpB-rbcL* digéré par l'enzyme Hae IIIa. 1,2,3, *C. arabica* (325, 211, 236); 4 *C. sp.X* (243); 5,6 *C. brevipes* (267, 276); 7,8, *C. canephora* (265,284); 9 *C. congensis* (291); 10 *C. eugenioides* (294); 11, 12, 13, 14 *C. liberica* (217, 257, 260, 303); 15 *C. millotii* (308); 16 *C. perrieri* (268); 17, 18 *C. pseudozanguebariae* (254, 255); 19 *C. racemosa* (253); 20, 21 *C. salvatrix* (250, 251); 22 *C. sp. Mayombe* (324); 23 *C. sp. Moloundou* (256); 24 *C. sp. N'koumbala* (275); 25 *C. sp. N'gongo II* (259); 26, 27 *C. stenophylla* (226, 232); 28 *P. manii* (247); 29 *P. ebracteolatus* (248).

#### Informations concernant le polymorphisme

Taq I et Rsa I ne révèlent pas de site de restriction polymorphe. La digestion du fragment *atpB-rbcL* par ces enzymes permet par contre de mettre en évidence des insertions/délétions courtes. Sur le fragment de 326pb obtenu après digestion par TaqI, il existe une délétion de 30 paires de bases environ chez *C. humblotiana* et une insertion de quelques paires de bases chez *C. racemosa*. L'ensemble des résultats obtenus sur ce fragment est présenté sur la figure 13.



**Figure 13.** Localisation du polymorphisme observé par RFLP sur le fragment intergénique *atpB-rbcL*

Les mutations détectées sont codées comme suit:

- I : site de restriction Hae III en position 380 sur la séquence (AGCC chez *C. arabica*, GGCC chez *C. pseudozanguebariae*),
- J : délétion d'environ 30pb entre le site RsaI 418 et le site TaqI 692,
- K : insertion d'environ 30pb entre les sites TaqI 370 et RsaI 418,
- L : insertion d'une dizaine de pb entre les sites RsaI 130 et TaqI 370.

Les caractères polymorphes mis en évidence sont codés pour chaque espèce dans le tableau 14.

espèce	I	J	K	L
<i>C.arabica</i>	0	0	0	0
<i>C.brevipes</i>	0	0	0	1
<i>C.canephora</i>	0	0	0	1
<i>C.congensis</i>	0	0	0	1
<i>C.eugenioides</i>	0	0	0	0
<i>C.humblotiana</i>	0	1	0	1
<i>C.liberica</i>	0	0	0	1
<i>C.millotii</i>	0	0	0	1
<i>C.perrieri</i>	0	0	0	0
<i>C.pervilleana</i>	0	?	0	1
<i>C.pseudozanguebariae</i>	1	0	0	1
<i>P.ebracteolatus</i>	0	0	0	1
<i>P.mannii</i>	0	0	0	1
<i>C.racemosa</i>	0	0	1	1
<i>C.salvatrix</i>	0	0	0	1
<i>C.sp Mayombe</i>	0	0	0	1
<i>C.sp Mouloundou</i>	0	0	0	0
<i>C.sp Ngongo 2</i>	0	0	0	1
<i>C.sp.X</i>	0	0	0	1
<i>C.stenophylla</i>	0	0	0	1

**Tableau 14.** Etats des 4 caractères polymorphes sur le fragment intergénique chloroplastique *atpB-rbcL* pour 20 taxons de caféiers.

Le nombre de caractères polymorphes mis en évidence est faible, bien que le fragment soit plus long (920 pb) que le fragment intergénique *trnL-trnF* (377pb) amplifié précédemment. Il existe donc probablement une hétérogénéité des taux de divergence entre les différentes séquences non codantes de l'ADNcp. La séquence de l'intergène *atpB-rbcL* pourrait avoir une vitesse d'évolution plus lente que celle d'autres intergènes. Ce résultat a également été observé entre 40 taxons de pommiers sauvages et cultivés, chez qui le séquençage complet de l'intergène *atpB-rbcL* a montré seulement 3 sites de mutation (Savolainen, communication personnelle). Ceci confirme que les séquences non codantes, comme les séquences codantes, ont des taux de substitution très variables (Olmstead & Palmer, 1994).

Chez les caféiers étant donné le faible polymorphisme de restriction observé pour la séquence de l'intergène *atpB-rbcL*, nous n'avons pas séquencé cette zone supposant que l'information phylogénétique qu'elle contient était très faible. Le polymorphisme codé au cours de cette étude sera retenu dans l'analyse phylogénétique de l'ensemble des données chloroplastiques.



## **5. PHYLOGENIE DES CAFEIERS A PARTIR DE DONNEES CHLOROPLASTIQUES**

Les résultats obtenus sur les variations de l'ADNcp de caféiers concernent d'une part le génome chloroplastique dans son entier, et d'autre part la divergence de 2 fragments intergéniques. Les caractères phylogénétiquement informatifs obtenus par chacune de ces approches prise indépendamment sont peu nombreux. L'analyse phylogénétique mérite donc d'être entreprise à partir de l'ensemble des données réunies sur l'ADNcp. C'est une analyse cladistique qui a été réalisée. C'est en effet la seule permettant l'analyse conjointe de données de diverses origines.

Nous avons, dans un premier temps, vérifié que les caractères des 2 études étaient homogènes afin d'éviter le biais que pourrait créer la non pondération (ou les erreurs de pondération) des caractères. Dans un second temps, la matrice regroupant l'ensemble des caractères variables de l'ADNcp chez les caféiers a été étudiée en parcimonie, et la phylogénie déduite discutée.

## **5.1. METHODOLOGIE**

### **5.1.1. Transformation des données**

Les données que nous avons accumulées concernant le polymorphisme de l'ADNcp sont relatives à des sites, des fragments de restriction ou des variations de séquences. L'analyse phylogénétique nécessite des caractères qui soient indépendants et codés de façon à pouvoir donner lieu à des analyses comparatives. Pour réaliser une analyse parcimonieuse non pondérée, il est également nécessaire de vérifier que chaque caractère est lié au même nombre d'évènements mutationnels. Dans notre cas, il s'agit de vérifier que la variation de chaque caractère est liée à une seule mutation sur la séquence.

Les caractères provenant des séquences sont de deux types : les mutations ponctuelles (transitions et transversions) qui ont été recodées arbitrairement (un état/0, l'autre état/1) en données binaires et les indels qui sont codés, quelle que soit leur longueur, comme une seule mutation. Le codage des mutations ponctuelles chez les individus possédant des délétions superposées est réalisé par l'introduction de points d'interrogation (figure 14). Ce cas s'est retrouvé à 14 reprises dans nos données. Ce codage permet la transcription d'événements emboîtés (mutations subséquentes à l'indel) en accord avec la logique de l'alignement qui est de mettre en évidence des mutations partagées, en minimisant le nombre de pas (Barriel, 1994).

Les mutations sur les séquences 1, 2 et 3 sont codées :				dell	mut1	mut2
1	AAAAAAAAAAAAAAAAAGAA			0	0	0
2	AAAAOoooooooooAAAAA			1	?	1
3	AAAAAAAAATAAAAAGAA			0	1	0
	D1	M1	M2			

**Figure 14.** Schématisation du codage de deux événements "emboîtés", par introduction de point d'interrogation.

Les caractères produits par les analyses RFLP sont également de 2 types. Le polymorphisme de longueur de fragments, du fait de la faible fréquence de ce type d'événement, sera considéré comme provoqué par des insertions ou délétions "uniques". Seuls trois caractères de ce type ont été codés, sur l'intergène *atpB-rbcL*. Les autres caractères que produit l'analyse RFLP sont les sites de restriction polymorphes. Pour les recoder en binaire comme des mutations ponctuelles, il faut que la mutation d'un site de restriction soit liée à un seul événement mutationnel.

La probabilité d'avoir "p" mutations ponctuelles sur un site de "k" bases (k=4 ou k=6), est:

$$P_p = C_p^k * x^p$$

où x est le taux de divergence entre les 2 séquences. La probabilité que la mutation d'un site soit due :

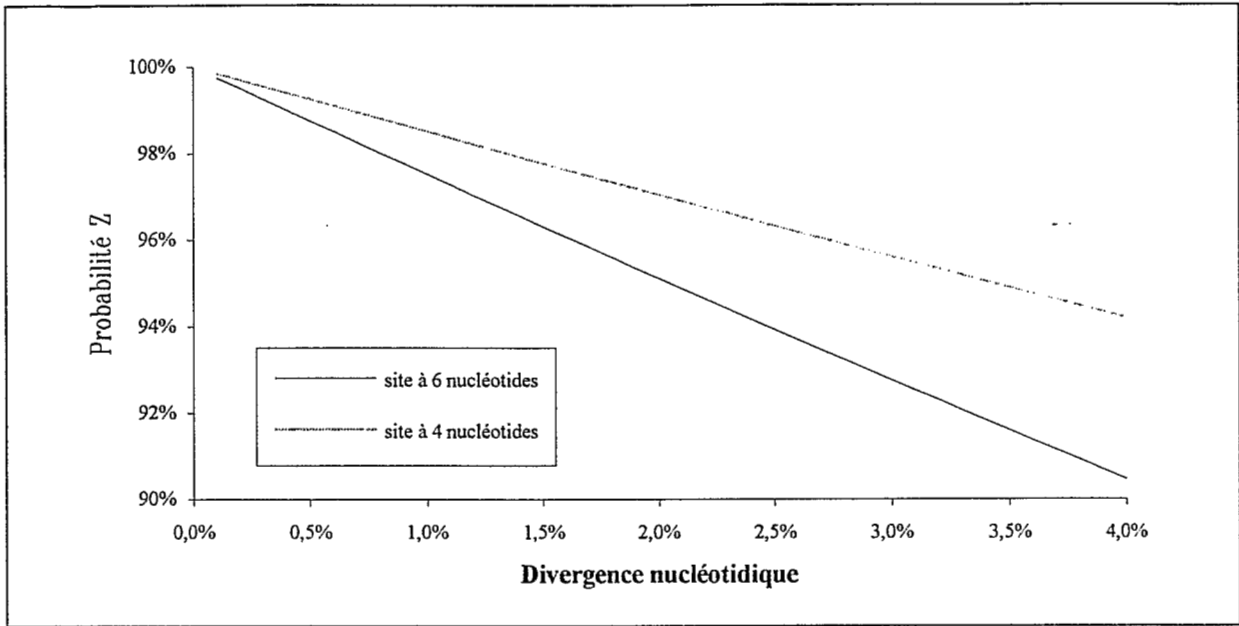
-à une seule mutation ponctuelle est donc,  $P_1 = k * x$ ,

-à plusieurs mutations ponctuelles est,  $P_{mul} = \sum_{p=2}^k [C_p^k * x^p]$

La probabilité (Z) que la mutation d'un site de restriction soit liée à une et une seule mutation ponctuelle est alors:

$$Z = (100 * P_1) / (P_1 + P_{mul}).$$

Cette probabilité a été calculée pour les sites produits par des enzymes ayant des sites de reconnaissance à 4 et à 6 paires de bases. On la représente en fonction de la divergence x (figure 15).



**Figure 15.** Probabilité **Z** pour que le changement d'état d'un site de restriction soit dû à une mutation ponctuelle en fonction de la divergence nucléotidique.

Chez les caféiers, nous observons une divergence nucléotidique maximale de 3% (sur le fragment intergénique, supposé le plus variable). Le changement d'état d'un site de restriction est donc provoqué par une seule mutation ponctuelle dans au minimum 93 % des cas pour une analyse réalisée avec des enzymes ayant des sites de reconnaissance à 6 nucléotides (96% pour sites à 4 nucléotides). Dans le cadre de notre étude, il est donc acceptable de considérer que la mutation d'un site de restriction est liée à une mutation unique.

### 5.1.2. Analyse phylogénétique

La reconstruction de la phylogénie de l'ensemble des espèces a été réalisée en employant la parcimonie de Wagner qui, comme nous l'avons précisée au chapitre 2, est la plus générale et n'impose aucune contrainte pour les changements d'état de caractères. Dans notre étude, c'est l'approche la plus cohérente puisque l'on ne connaît pas le mode d'évolution des caractères.

Le choix de l'"outgroup" peut largement affecter les chances d'obtenir une topologie correcte (Smith, 1994). Il semble que la meilleure façon de réduire ce biais soit d'utiliser un "outgroup" proche, et qui ait peu divergé par rapport aux positions invariantes du groupe à étudier (Smith, 1994). Etant donné la très faible divergence chez les caféiers, nous avons choisi *Gardenia grandiflora* qui semble posséder les qualités requises.

La matrice de données a donc, dans un premier temps, été analysée à l'aide du logiciel MIX du groupe de programme "PHYLIP" (Felsenstein, 1989) en utilisant le *Gardenia grandiflora* comme "outgroup". Sur l'arbre phylogénétique les mutations autapo- et synapomorphiques sont signalées sur chaque branche. Les qualités (rapport "signal/bruit", "validité") de l'arbre sont estimées grâce au calcul de l'index de consistance (Farris, 1989), au nombre d'arbres de parcimonie maximum obtenus et au "biais" lié à la taille de la matrice (Hillis & Huelsenbeck, 1992).

Dans un deuxième temps, une analyse par rééchantillonnage des caractères (de type "bootstrap"), visant à évaluer la robustesse de cette phylogénie (Felsenstein, 1994), a été réalisée à l'aide du logiciel "BOOT" du groupe de programme "PHYLIP" (Felsenstein, 1989). Cette démarche a été appliquée à plusieurs reprises (avec introduction des espèces dans un ordre différent, sur plusieurs jeux réduits d'espèces du sous-genre *Coffea*), au cours d'analyses de 100 réplifications. La robustesse des arbres phylogénétiques qui en sont déduits, est estimée en fonction de la valeur de "bootstrap" qui supporte chaque branche. Dans la pratique, c'est l'arbre de consensus majoritaire à 50% qui est présenté (toute valeur de "bootstrap" inférieure à 50% étant représentée sous forme de polytonie).

## **5.2. RESULTATS**

### **5.2.1. Bilan des caractères retenus**

L'ensemble des caractères étudiés ont été rassemblés dans une matrice de type "individus X caractères".

La combinaison des trois études nous a permis d'étudier 112 sites de restriction (dans l'analyse RFLP de l'ADNcp total), 377 positions nucléotidiques (étude de la séquence intergénique trnL-trnF) et 24 fragments de restriction (analyse RFLP du fragment intergénique atpB-rbcL), soit un total de 513 descripteurs. Ces descripteurs ne sont pas tous polymorphes entre les génotypes étudiés. Dans le cas de la séquence qui est alignée sur 377 positions, le polymorphisme ne porte que sur 35 mutations ponctuelles et 5 délétions.

Le nombre des descripteurs est ramené de 513 à 55 caractères. Chacun de ces caractères (produit par les diverses approches) est assimilé à un événement mutationnel unique.

L'ensemble des individus étudiés n'est pas discriminé. En effet, plusieurs génotypes appartenant pourtant à des espèces différentes possèdent une suite d'états de caractères complètement identiques. Le nombre "d'objets" distincts est de cette façon réduit à 26 OTU<sup>1</sup>.

La matrice corrigée contient donc 26 OTU's X 55 caractères. Elle est présentée dans la figure 16.

---

<sup>1</sup>pour "Operational Taxonomic Unit"

[illegible]

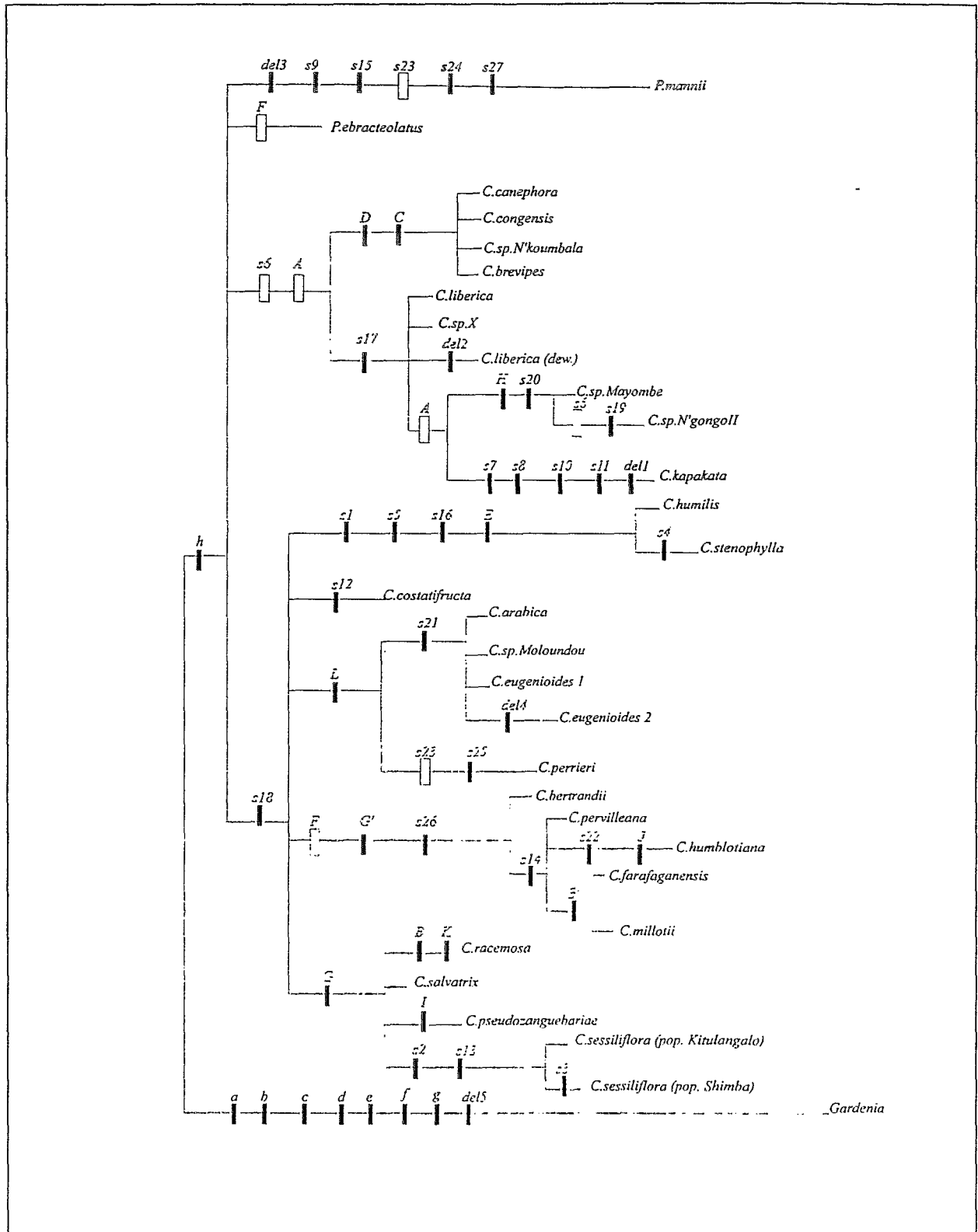
Figure 16. Matrice corrigée rassemblant l'ensemble des caractères présentant du polymorphisme (issus de 3 approches différentes) sur l'ADNcp des caféiers.

### 5.2.2. Parcimonie de Wagner

Cette analyse permet de construire plus de 100 arbres également parcimonieux, mais il existe un seul arbre de consensus strict (Darlu & Tassy, 1993) comportant plusieurs polytonies. Il est présenté sur la figure 17. Chaque changement d'état d'un caractère est représenté par une barre verticale. Par convention, les barres pleines représentent les caractères acquis sur un seul embranchement et qui ne réversent pas (non-homoplasiques). Les barres blanches figurent les caractères homoplasiques. Chaque caractère repéré par une lettre ou un numéro correspondant au type d'analyse dont il provient, est identifié dans la matrice corrigée (figure 16).

L'arbre de consensus strict a une longueur de 58 pas, dont 4 sont dus à de l'homoplasie, 2 caractères provenant des données de séquence et 2 provenant des données RFLP. L'indice de cohérence calculé à partir de l'ensemble des caractères est de 0,93 (Farris, 1989). Cet indice est élevé ce qui signifie que l'homoplasie est faible. L'examen de la distribution des mutations sur les branches permet de constater que cet indice de cohérence élevé cache un taux d'autapomorphie élevé. En effet, si l'on considère chaque objet terminal ou feuille comme une seule "espèce" au sens phylogénétique du terme, le rapport de synapomorphies / autapomorphies est de : 20/38. Ce faible taux de synapomorphie est responsable du fait que certaines branches ne soient supportées que par 1 ou 2 mutations, voire aucune. C'est ce phénomène caractéristique qui est à l'origine de l'obtention d'un nombre élevé d'arbres (strictement dichotomiques) également parcimonieux. Certains "embranchements" ne sont donc pas résolus (présence de polytonies), mais l'arbre de consensus strict possède une structuration suffisante pour permettre de proposer des hypothèses phylogénétiques.



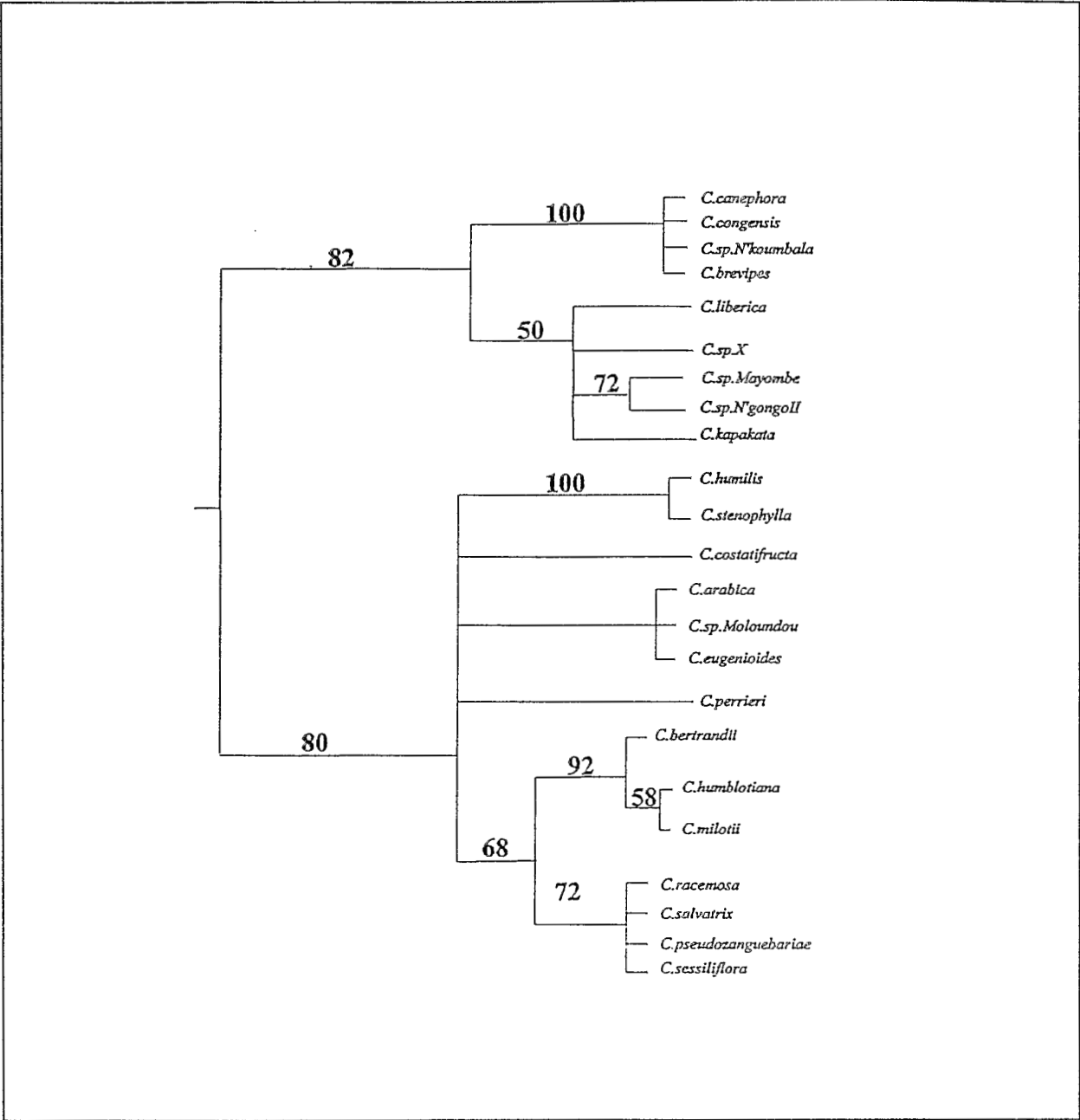


**Figure 17.** Arbre phylogénétique de parcimonie maximum (consensus strict) obtenu à partir de données chloroplastiques. Les "barres" verticales représentent les mutations supportant chaque clade. Par convention, les barres pleines représentent les caractères non-homoplasiques.

### 5.2.3. "Bootstrap"

La topologie de l'arbre de parcimonie et la taille de la matrice nous ont incité à utiliser une matrice de taille réduite de façon à diminuer le temps de calcul nécessaire à 100 rééchantillonnages. Les 2 espèces malgaches *C.farafaganensis* et *C.pervilleana* ont été exclues de la matrice parce qu'elles comportaient trop de caractères non déterminés (séquences non obtenues). Pour l'espèce *C.sessiliflora*, un seul génotype a été retenu (les 2 génotypes ne différaient que par une mutation). Enfin, les 2 espèces appartenant au genre *Psilanthus* ont été volontairement écartées, afin de réaliser une analyse à partir des seules espèces du sous-genre *Coffea*. L'arbre de consensus majoritaire à 50% est présenté sur la figure 18. Six branches seulement sont supportées par des valeurs de "bootstrap" supérieures ou égales à 80.

Si les données ne contiennent pas d'homoplasie, le fait qu'une branche soit supportée par au moins 3 synapomorphies semble suffisant pour que la valeur de "bootstrap" qui lui est associée soit supérieure à 95% (Felsenstein, 1994). C'est effectivement ce que l'on observe sur l'embranchement *C.stenophylla* et *C.humilis* (valeur de bootstrap=100). Des valeurs de "bootstrap" élevées (proches de 100) ne sont donc pas forcément significatives d'une grande longueur de branche (nombre de synapomorphies élevées).



**Figure 18.** Arbre de "bootstrap" (consensus majoritaire à 50%) obtenu à partir de données chloroplastiques. Les chiffres indiqués sur les branches sont les valeurs de "bootstrap" obtenues pour l'embranchement considéré.

Lorsqu'il existe trop peu de synapomorphies (moins de trois par branche) on observe des valeurs de "bootstrap" faibles. C'est ce que l'on observe sur la majorité des clades de la phylogénie des caféiers. Ces faibles valeurs de bootstrap, lorsqu'elles ne sont pas liées à de l'homoplasie, sont liées au manque de structuration de type dichotomique. Des valeurs de

"bootstrap" inférieures à 80 peuvent donc être le signe d'un réel phénomène évolutif et suggérer une radiation rapide.

### **5.3. DISCUSSION**

#### **5.3.1. Limites de l'étude**

##### Matrice analysée

Il est généralement convenu que pour assurer la robustesse d'un arbre phylogénétique, chaque clade doit être supporté en moyenne par 5 caractères (Gascuel, communication personnelle).

En acceptant cette hypothèse et pour un nombre identique d'objets, la taille de la matrice nécessiterait la multiplication par 3 du nombre de caractères informatifs. De plus cette matrice révèle après analyse un rapport de synapomorphies/autapomorphies faible qui se traduit par un faible taux de mutations informatives. L'augmentation du nombre de caractères informatifs paraît donc difficile.

Une autre possibilité consiste à limiter le nombre d'objets. Or la position relative d'espèces jamais étudiées dans le passé demeure un des objectifs de notre travail. Nous ne souhaitons donc pas réduire le nombre d'espèces étudiées.

##### Phylogénie chloroplastique

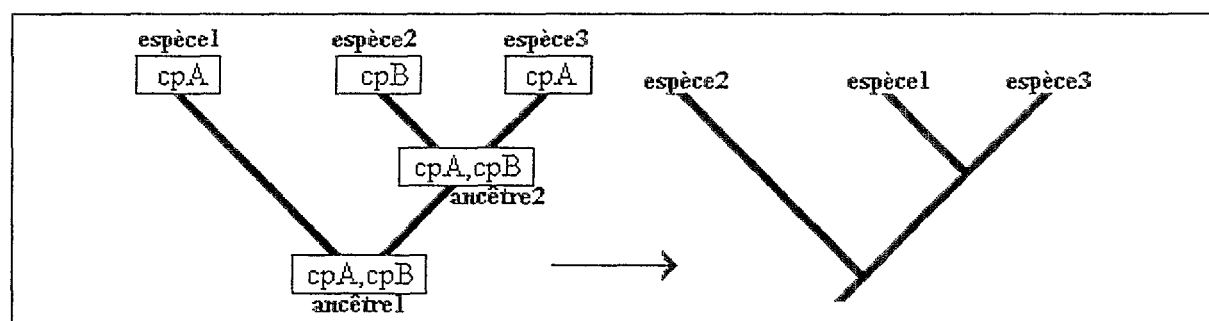
Les feuilles (ou noeuds terminaux) d'un arbre phylogénétique construit à partir de données provenant de l'ADNcp (RFLP ou séquences) ne sont pas les espèces porteuses de ces ADNcp mais bien les génomes (voire les séquences de gènes) chloroplastiques. Ces données peuvent produire un arbre phylogénétique très robuste, mais différent de l'arbre des espèces. Les phylogénies établies à partir de telles données ne sont donc pas nécessairement les phylogénies des espèces (Doyle, 1992). Par contre la phylogénie des espèces peut être estimée à partir de telles phylogénies. Il existe parfois des incohérences entre phylogénie construite à partir de données morphologiques d'une part, et moléculaires d'autre part. Ces incohérences, pour l'aspect phylogénie moléculaire, sont principalement dues à trois phénomènes : les

introgressions, la duplication de gènes et le phénomène de "lineage sorting" (voir Figure 19) (Doyle, 1992).

Les introgressions consistent en un transfert de gènes à travers la barrière reproductive. Dans des cas extrêmes, il peut s'agir d'un chlorotype extérieur (b) qui remplace le chlorotype initial (a) d'une plante (A). Si une phylogénie est construite à partir de données chloroplastiques, la plante hôte (A) va paraître phylogénétiquement proche de la plante ayant "donné" son type de chloroplaste (B), bien que leur histoire évolutive soit différente. Ce phénomène semble assez rare bien qu'il puisse exister entre espèces ayant divergé depuis plus de 15 à 20 millions d'années (Rieseberg & Soltis, 1991).

La duplication de gènes ou les familles multigéniques peuvent également induire des reconstructions phylogénétiques erronées. Le mélange de séquences orthologues (produites par spéciation) et paralogues (produites par duplication) induit une reconstruction phylogénétique qui, si elle reflète l'histoire du gène, ne reflète pas l'arbre des espèces. Ce phénomène est peu probable dans le cas de l'ADNcp, mais il faut en tenir compte pour l'étude de l'ADN ribosomique (par exemple).

Le "lineage sorting" est un concept lié à l'existence de polymorphisme intraspécifique (voire intragénotype) de l'ADNcp chez les individus ancêtres (figure 19). Bien que cela ait été souvent ignoré, presque toutes les espèces chez qui plusieurs individus par espèces ont été échantillonnés ont montré un polymorphisme intraspécifique de l'ADNcp.



**Figure 19.** "Lineage sorting". Un ancêtre possédant un ADNcp polymorphe (cpA et cpB) donne naissance à trois espèces selon le schéma de gauche. La reconstruction phylogénétique (quelle que soit la méthode employée) va rapprocher les espèces 1 et 3, induisant un embranchement incorrect.

Bien que les caractères moléculaires soient des caractères phylogénétiques possédant de nombreuses qualités par rapport aux données morphologiques, ce ne sont donc pas des données idéales. Pour éviter les conclusions phylogénétiques erronées, basées sur des données concernant l'ADNcp, plusieurs approches sont envisagées. Tout d'abord la comparaison avec des hypothèses phylogénétiques basées sur des gènes nucléaires, voire des données morphologiques. On peut également envisager une stratégie d'analyse modifiée, en étudiant

séparément plusieurs jeux de données chloroplastiques (2 types de séquences), suivie de leur analyse conjointe (Rieseberg & Soltis, 1991).

Les comparaisons de phylogénies basées sur des caractères morphologiques et des données moléculaires se multiplient. Dans une comparaison de ce type pour 36 espèces d'anémones et 7 genres voisins, les clades terminaux sont quasi-identiques (Hoot *et al.*, 1994). Par contre il existe des incohérences dans les clades plus anciens. L'arbre phylogénétique basé sur des données chloroplastiques montre une meilleure résolution à haut niveau taxonomique, accompagné de moins d'homoplasie et de valeurs de "bootstrap" plus élevées. La phylogénie qui en est déduite est donc considérée comme meilleure. A bas niveau taxonomique (Freudenstein & Doyle, 1994) on a montré chez 35 populations de *Corallorhiza maculata* (Orchidaceae) qu'il existait une corrélation forte entre le type d'ADNcp et la morphologie.

### 5.3.2. Type de spéciation

#### Topologie globale

L'arbre phylogénétique possède une forme partiellement "en râteau", signe que l'évolution ne peut pas être représentée de façon strictement dichotomique. Cette topologie n'est pas provoquée par de l'homoplasie mais elle est associée à un faible taux de synapomorphies/autapomorphies. Du point de vue théorique, on peut se demander s'il existe un arbre parfaitement dichotomique lorsqu'il existe de l'imprécision liée à la réticulation autour de l'embranchement. Ce problème (qui se posait par exemple pour le clade Homme/Chimpanzé/Gorille) peut ne pas être résolu malgré une accumulation considérable de caractères. Dans de tels cas, la topologie en râteau n'est pas un artefact lié aux données étudiées, mais bien le signe d'une explosion radiative rapide. La véritable solution ne se retrouve donc pas dans une résolution à tout prix des embranchements polytoniques, mais plutôt dans notre capacité à proposer un modèle de l'évolution des espèces considérées.

#### Place des *Psilanthus*

Un certain nombre d'éléments peuvent tout d'abord être évoqués concernant la position des deux échantillons appartenant au genre *Psilanthus* : bien que très proches des autres espèces de caféiers, les deux espèces de *Psilanthus* sont dérivées directement du premier embranchement (Figure 17). Les espèces du sous-genre *Coffea* appartiennent probablement au même phylum que les espèces du genre *Psilanthus*. De plus, aucune synapomorphie ne

supporte un éventuel phylum composé des espèces du genre *Psilanthus*. Aucun argument ne donne à ces dernières le statut de sous groupe monophylétique.

### Datation

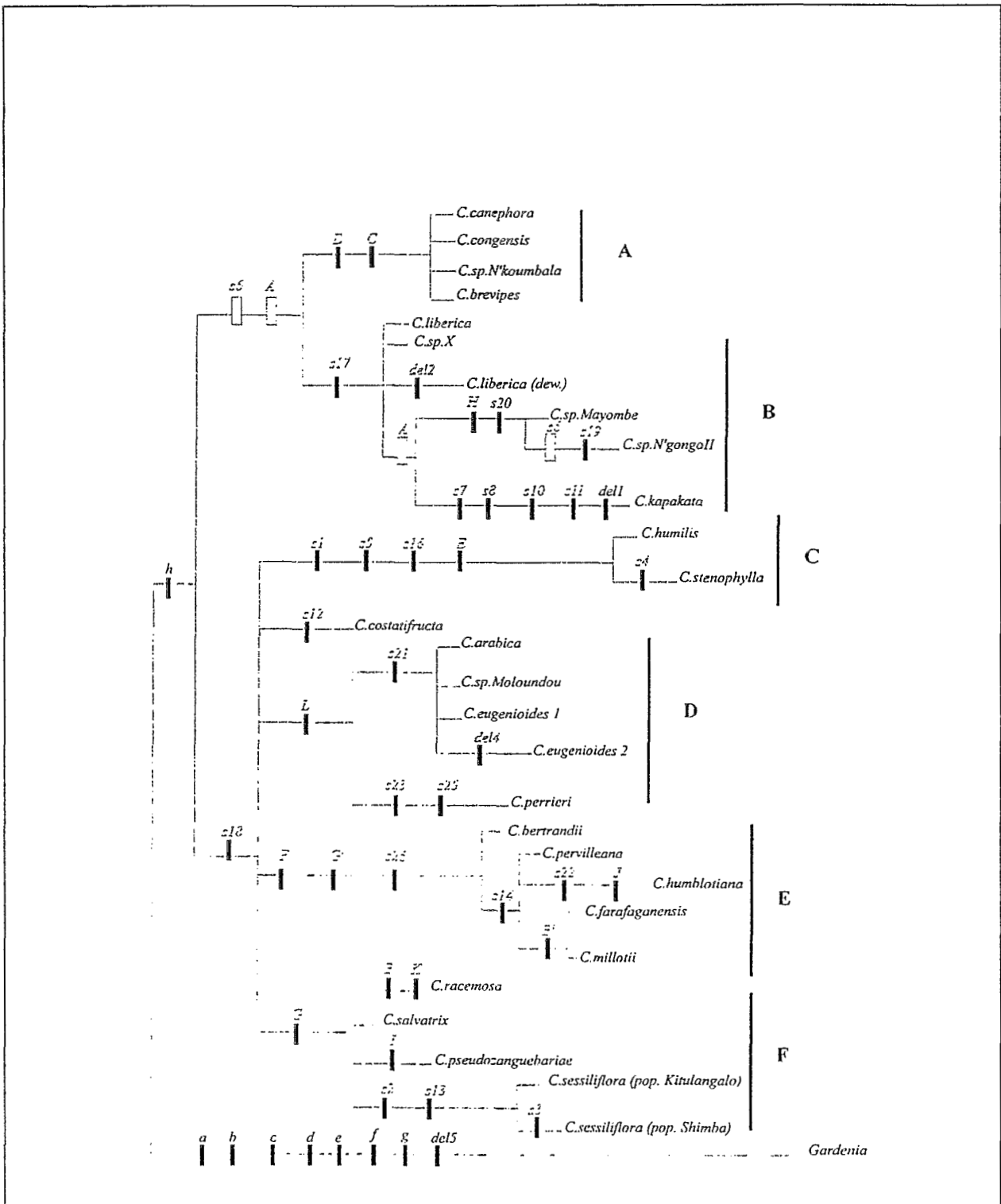
Une hypothèse qui pourrait permettre de dater l'embranchement des caféiers est celle de Leroy qui proposait un isolement primitif des caféiers malgaches au cours de la dislocation gondwanienne, il y a environ 80 Millions d'années (Leroy, 1982). La topologie globale de l'arbre, la présence d'une espèce malgache (*C.perrieri*) sur un clade supportant des espèces est-africaines et surtout la position du clade regroupant la plupart des espèces malgaches, ne confirme pas cet isolement primitif. On n'a donc aucun moyen de poser d'hypothèse concernant l'horloge moléculaire.

Le taux de substitution connu pour les zones non codante semble toutefois assez peu variable, ce qui pourrait permettre de proposer une date pour la première explosion radiative du genre *Coffea*. Le taux de substitution sur la séquence (codante) de *rbcL* varie de 0,1 à 2. 10<sup>-10</sup> substitutions/site/an (Chase *et al.*, 1993; Clegg, 1993), et jusqu'à 6. 10<sup>-10</sup> selon certains auteurs (Doebley *et al.*, 1990; Frascaria *et al.*, 1993). Sur les séquences non codantes, le taux de substitution nucléotidique est de l'ordre de 3,2 à 5,7. 10<sup>-10</sup> substitutions/site/an pour les séquences intergéniques ribosomiques ITS1 et ITS2 chez les Winteraceae (Suh *et al.*, 1993), ce qui est relativement faible. Sur l'ADNcp, les données concernant l'intergène *atpB-rbcL* donnent des valeurs de ce taux de substitution allant de 11 à 19. 10<sup>-10</sup> substitutions/site/an en fonction de la lignée sur laquelle il a été estimé (Kawagoe & Kikuta, 1991; Zurawski & Clegg, 1987). Bien que plusieurs travaux aient été réalisés sur l'intergène *trnL-trnF*, le taux de substitution (par site/an) n'a pas été estimé. En admettant une fourchette moyenne de 5 à 15 10<sup>-10</sup> substitutions/site/an pour l'ensemble des caractères que nous avons étudiés (Wolfe *et al.*, 1987), cela signifierait que la non résolution d'un embranchement (topologie en râteau) est le produit d'une explosion radiative qui s'est produite en moins de 4Ma. Sur cette base, qui constitue seulement un ordre de grandeur, on peut proposer (à partir de la matrice de données et de la position des mutations sur l'arbre de parcimonie) une estimation de la première différenciation du genre *Coffea* entre 10 et 30Ma BP. Cette indication ne confirmerait pas la proposition de Leroy d'une existence précoce du genre *Coffea* aux environs de 100Ma BP (Leroy, 1982). Ces données supporteraient plutôt l'hypothèse d'une première différenciation des caféiers au moment de la formation des grands rifts africains à environ -20Ma BP.

Organisation en clades

Aucune mutation ne permet de supporter la monophylie des caféiers dans leur ensemble ou du genre *Coffea* par rapport au genre *Psilanthus*. Toutefois, si l'on considère exclusivement les espèces du sous-genre *Coffea*, on peut admettre une différenciation primitive en 2 clades qui se sont chacun très rapidement subdivisés en 6 clades. Chacun de ces clades correspondrait à un chlorotype ancêtre. Seule l'espèce *C.costatifructa* n'est pas classée, sans qu'il soit possible de savoir si c'est parce que cette espèce préfigure la différenciation d'un nouveau clade ou parce que le nombre de caractères liés à chaque embranchement est insuffisant pour l'associer à l'un des clades précédents (figure 20).





**Figure 20.** Arbre phylogénétique (parcimonie maximum, consensus strict) des espèces du sous-genre *Coffea*. Les "barres" verticales représentent les mutations supportant chaque clade (les barres pleines représentent les caractères non-homoplasiques). Les 6 clades déterminés sont repérés par les lettres A,B,C,D,E,F.

Le premier embranchement supporte deux clades (A et B). L'un (A) est très "homogène" puisqu'il est supporté par 2 mutations, et que les 4 espèces qui le composent ne portent pas d'autapomorphie. Le second regroupe des espèces beaucoup plus hétérogènes. Les espèces que l'on retrouve dans ces 2 clades sont originaires de toute la zone guinéo-congolaise.

Le deuxième embranchement supporte au moins 4 clades. Un premier (C) rassemble 2 espèces, *C.stenophylla* et *C.humilis*, occupant des régions voisines (Côte d'Ivoire jusqu'à la frontière guinéenne) mais possédant des phénotypes très différents. Ce clade est supporté par 4 mutations et une valeur de "bootstrap" élevée, qui sont des indices de la monophylie de ce groupe. Deux clades (E et F) possèdent sans conteste un ancêtre commun. En effet alors que les deux caractères B et G issus de l'étude RFLP supportent le clade F, les "caractères" B' et G' (signes de l'hétérogénéité des caractères B et G) supportent le clade E. Ceci peut s'expliquer par l'existence d'un ancêtre commun portant de l'ADNcp hétérogène pour les caractères B et G, et la transmission à la descendance de l'un (0), l'autre (1) ou les deux (0/1) caractères. Enfin le clade D réunit 3 espèces très proches (*C.arabica*, *C.eugenoides* et *C.sp.Moloundou*) ainsi qu'une espèce de *Mascarocoffea*, *C.perrieri*.

### 5.3.3. Comparaison avec des données concernant la diversité

#### Comparaison avec des données agromorphologiques

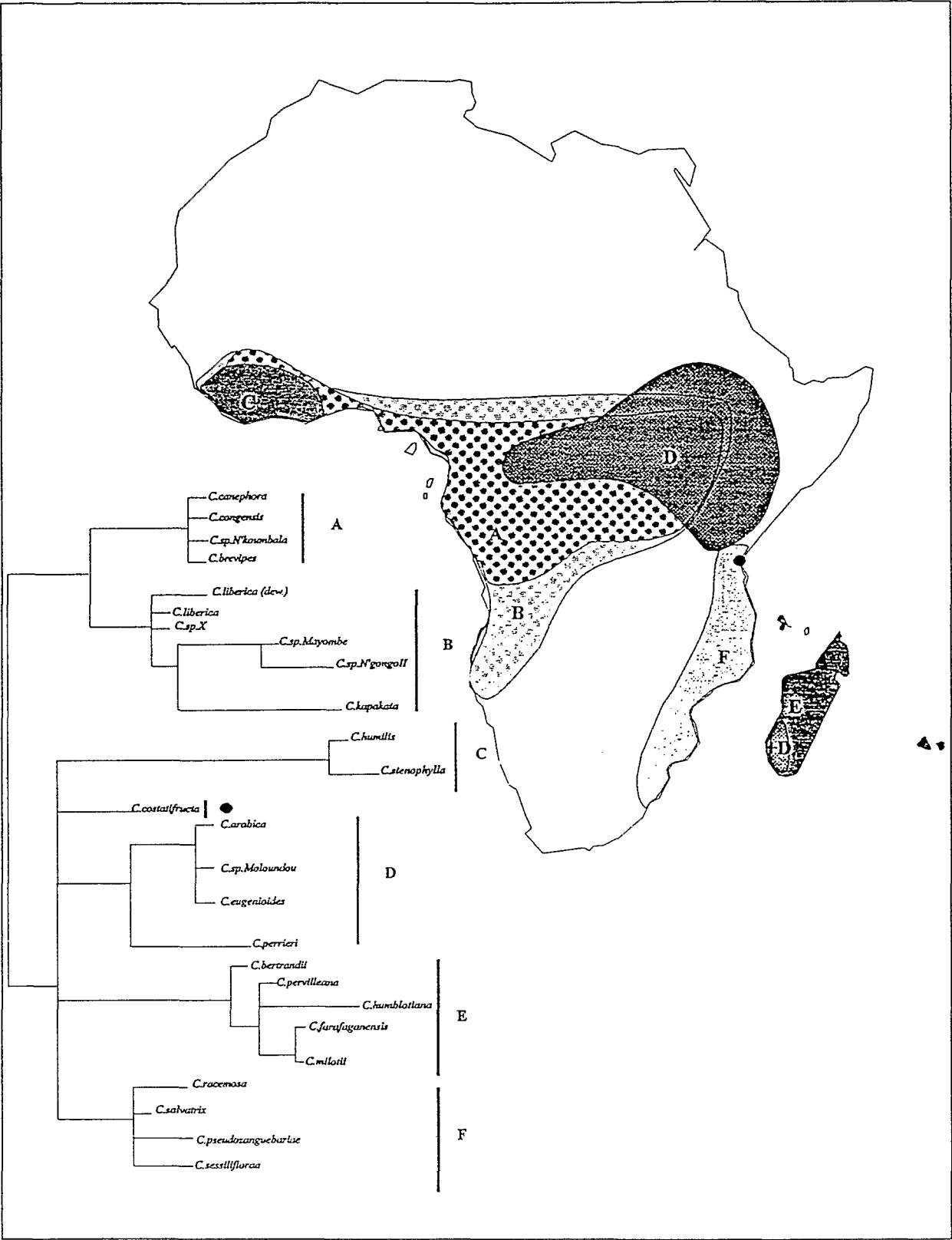
L'organisation en clade ne semble pas pouvoir (dans la majorité des cas) être corrélée à une proximité phénotypique ou une ressemblance phénologique. Le clade C regroupe très fortement 2 espèces (*C.stenophylla* et *C.humilis*), bien qu'aucun caractère morphologique ou phénotypique n'ait pu laisser présumer ce résultat. Au contraire, certaines caractéristiques originales de *C.stenophylla* (couleur des fruits) et la morphologie très particulière de *C.humilis*, semblaient les isoler particulièrement. De même le clade E regroupe la majeure partie des *Mascarocoffea* qui sont morphologiquement très hétérogènes, mais n'inclut pas *C.perrieri*. Or *C.perrieri* (clade D) appartient comme *C.bertrandi* (clade E) à la série botanique des multiflorae qui regroupe des espèces ayant des caractéristiques morphologiques communes (Charrier, 1976). Le cas du clade D est particulier parce qu'il regroupe des espèces ayant des particularités morphologiques et génétiques intéressantes. *C.arabica* et *C.eugenoides* sont deux taxons afro-montagnards dont les relations de parenté avaient déjà été envisagées. *C.sp.Moloundou*, par contre est une espèce récemment découverte, la seule espèce diploïde qui comme *C.arabica* (tétraploïde) soit autogame.

Au contraire, les 4 taxons réunis dans le clade F (*C.pseudozanguebariae*, *C.racemosa*, *C.salvatrix* et les 2 populations de *C.sessiliflora*), bien que leur caractéristiques biochimiques ne soient pas homogènes (Anthony *et al.*, 1993), présentent un certain nombre de caractéristiques morphologiques et phénotypiques communes (Anthony, 1992). Cette homogénéité de comportement (petite feuilles épaisses, cycle de fructification court) est en partie liée à leur adaptation à une sécheresse relative.

L'ensemble de ces caractéristiques, associé à l'important polymorphisme intraspécifique pour ces caractères, semble suggérer que l'adaptation morphologique et phénologique sont des phénomènes le plus souvent indépendants des relations de parenté entre espèces.

#### Comparaison avec la distribution géographique

La figure 21 représente la répartition actuelle des espèces constituant chaque clade. Les trois clades A,B et C sont situés strictement à l'ouest de la dorsale du Kivu. Au moins un (mais deux si l'on considère que l'espèce *C.costatifructa* préfigure un nouveau clade), le clade F, est situé strictement à l'est. Le clade E regroupe des espèces strictement endémiques de Madagascar et des îles Mascareignes. La distribution actuelle des espèces du clade D est par contre à cheval sur les deux zones africaines.



**Figure 21.** Répartition géographique des 6 clades de caféiers définis par l'analyse des variations de l'ADNcp.

### 5.3.4. Eléments concernant la spéciation des caféiers

#### Origine de *C.arabica*

Le cas du clade D est particulièrement intéressant, en premier lieu à cause de sa répartition géographique et parce qu'il englobe l'espèce *C.arabica*, la seule espèce tétraploïde. Ce clade regroupe fortement trois espèces originaires de régions écologiquement différentes (*C.arabica*, *C.eugenoides*, *C.sp.Moloundou*). *C.sp.Moloundou* est un taxon provenant de la forêt de basse altitude et dont plusieurs populations ont été récoltées de part et d'autre de la frontière Cameroun-Congo. L'affinité et la possible parenté entre *C.sp.Moloundou* et *C.arabica* n'avait jamais été évoquée, bien que ces 2 espèces soient les 2 seuls taxons sauvages à avoir comme mode de reproduction l'autogamie préférentielle. *C.eugenoides* par contre, est, comme *C.arabica* une espèce afro-montagnarde (Anthony, 1992). Sa possible parenté (maternelle) avec *C.arabica* avait déjà été évoquée (Berthou *et al.*, 1983). Son comportement original en croisement la place en position intermédiaire pour tout regroupement entre espèces diploïdes (Louarn, 1993).

L'espèce *C.arabica* a été le plus souvent isolée, à cause de sa tétraploïdie et de caractères particuliers (autogamie). Les résultats obtenus permettent d'affirmer que cette espèce a une histoire liée de façon très étroite à celle de deux espèces diploïdes d'Afrique centrale et de l'est. L'étude de l'ADNcp (à cause de son hérédité monoparentale) ne permet pas de préciser l'origine auto- ou allotétraploïde de cette espèce. Par contre il est possible d'affirmer que l'histoire de *C.arabica* est très récente par rapport à celle de l'ensemble des caféiers. En effet aucune mutation ne permet de distinguer son chlorotype de celui des 2 autres espèces. Si les valeurs de l'horloge moléculaire précédemment décrite sont respectées, alors la spéciation de *C.arabica* a eu lieu au cours du quaternaire.

#### Relations entre espèces malgaches et est-africaines

En appliquant la valeur moyenne de l'horloge moléculaire définie ci-dessus, la séparation des caféiers malgaches et africains daterait de 5 à 20Ma BP soit largement après la séparation de l'Afrique et de Madagascar. Pour cela il faut admettre qu'un passage ait pu se faire entre le continent et la Grande île. Les affinités les plus importantes retrouvées entre la flore africaine et la flore malgache concernent les bryophytes et autres cryptogames (Pocs, 1975). Ceci est lié selon l'auteur à l'antériorité de l'embranchement des bryophytes, par rapport aux phanérogames, et donc à la possibilité pour ces espèces de coloniser "directement" Madagascar (Pocs, 1975). Pourtant Raven et Axelrod suggèrent que bien que les continents aient été séparés, la migration des plantes a toujours été possible via les îles volcaniques. Certaines

plantes et certains animaux ont apparemment traversé cette "barrière" d'eau bien après le milieu du crétacé (Raven & Axelrod, 1974).

A l'exception du cas particulier de *C.perrieri*, qui mérite d'être analysé à part, l'examen des positions des mutations sur les branches de l'arbre permet de constater 2 phénomènes caractéristiques. Les mutations B et G sont acquises sur le clade F alors que les mutations B' et G' (signes de profils hétérogènes) sont acquises exclusivement sur le clade E. Ce résultat pourrait être lié au phénomène de "lineage sorting". La présence de ces caractères hétérogènes sur un seul embranchement peut signifier qu'il y a une histoire plus récente commune entre lignées de part et d'autre du canal du Mozambique. La possibilité de relations privilégiées entre espèces est-africaines et malgaches a été suggérée encore récemment à partir de données biochimiques (Anthony, 1992; Rakotomalala, 1992).

Le cas de *C.perrieri* est plus particulier. D'un point de vue taxonomique *C.perrieri* fait partie de la série des Multiflorae, comme *C.bertrandii*. Cette série semble pourtant assez hétérogène. Elle recouvre une grande diversité d'adaptations altitudinales, pédologiques et climatiques (Charrier, 1976). L'espèce *C.perrieri* possède certains caractères agronomiques intéressants (vigueur) mais également une amertume caractéristique. Elle pourrait posséder une origine différente des autres Mascarocoffea. Selon nos résultats, cette espèce aurait colonisé Madagascar plus récemment que les espèces du clade E. L'originalité de ce taxon *C.perrieri* a déjà été signalée pour des caractéristiques génétiques (Charrier, communication personnelle) et biochimiques (Rakotomalala, 1992).

#### Spéciation très récente de certaines espèces

Alors que l'on observe un taux d'autapomorphie particulièrement élevé, certains taxons ne se distinguent par aucune autapomorphie (clade A, clade D). C'est ainsi que les unités taxonomiques, définies d'un point de vue phylogénétique, regroupent parfois plusieurs unités taxonomiques. Ce résultat est particulièrement net dans le cas du clade A, réunissant les 4 taxons : *C.canephora*, *C.congensis*, *C.brevipes*, *C.sp.N'koumbala*. La fertilité hybride est très élevée entre *C.canephora* et *C.congensis* (Louarn, 1993), et suggérerait une proximité entre ces deux espèces. Aucun résultat concernant la fertilité hybride des deux autres espèces n'est disponible du fait de leur introduction récente en croisement. Pourtant ces espèces sont clairement définies pour leurs caractéristiques botaniques, biochimiques et isozymiques. On peut donc suggérer que ces espèces, différenciées du point de vue taxonomique, appartiennent à la même espèce au sens phylogénétique du terme.

#### 5.4. BILAN

Les données accumulées sur l'ADNcp des caféiers révèlent 55 caractères polymorphes entre 26 OTU's. Certains individus, et même certaines espèces (*C.canephora*, *C.congensis*, *C.brevipes*, *C.sp.N'koumbala*) ne sont pas discriminés. Au contraire, trois espèces (*C.liberica*, *C.eugenoides* et *C.sessiliflora*) montrent de la variation intraspécifique pour un seul caractère.

La topologie de la phylogénie obtenue est partiellement en râteau mais il est possible de distinguer au moins 6 clades correspondant à des espèces qui pourraient avoir un génome chloroplastique ancêtre commun. Ces 6 groupes réunissent des espèces contemporaines ayant des zones d'endémisme strictement limitées à une partie de la zone d'occupation globale des caféiers. L'appartenance à un clade semble donc corrélée à une origine géographique.

Si l'on applique une valeur moyenne d'horloge moléculaire (5 à 15.  $10^{-10}$  substitutions/site/an) on peut estimer la première différenciation des espèces entre 10 et 30 MaBP. La différenciation de certaines espèces actuelles est au contraire très récente, par rapport à l'histoire des caféiers.

Chez certains génotypes appartenant aux clades E et F, il existe au moins 2 types de copies de l'ADNcp. Ce phénomène qui est rarement décrit semble relativement fréquent chez les caféiers. Il pourrait indiquer une ascendance commune des espèces situées de part et d'autre du canal du Mozambique. Il est pour cela nécessaire de postuler que l'ancêtre portait un ADNcp hétérogène qu'il aurait transmis différenciellement par un phénomène de "lineage sorting".

## **6. VARIATION DE LA TAILLE DU GENOME NUCLEAIRE DES CAFEIERS ET RELATIONS ENTRE ESPECES**



Toutes les espèces de caféiers que nous avons étudiées sont diploïdes avec  $2n=22$  chromosomes, à l'exception de *C.arabica* (qui est tétraploïde). Ces espèces présentent toutefois une très grande diversité, liée en partie à leur origine phylogénétique mais également à des adaptations à des écosystèmes différents. Pour tenter d'amorcer une interprétation plus générale des relations entre espèces, et en complément des données acquises par ailleurs, nous avons déterminé la taille des génomes nucléaires des espèces de caféiers. De plus, Il nous a semblé intéressant de vérifier s'il existait une variation et si elle était génotype spécifique, espèce spécifique ou groupe d'espèces spécifique (Cros *et al.*, 1994). La mise en évidence de variations de taille du génome nucléaire a ensuite été comparée à l'appartenance à un groupe d'espèces ayant une origine commune (même clade).

La taille physique du génome est directement dépendante de la quantité d'ADN par noyau : il existe en effet une relation linéaire entre la quantité d'ADN en picogrammes et la longueur du génome en paires de bases (Price, 1988). Pour quantifier la quantité d'ADN par noyau, nous avons utilisé la technique de cytofluorométrie en flux (CMF) (Dolezel *et al.*, 1989; Arumuganathan & Earle, 1991; Dolezel, 1991). Elle est couramment employée depuis le milieu des années 80 dans le domaine végétal pour estimer les variations de niveau de ploïdie (Brown *et al.*, 1991). La CMF permet de mesurer pour chaque noyau isolé (séparé dans un flux de liquide) la fluorescence émise après excitation par un faisceau laser. Elle s'avère beaucoup plus rapide que la microspectrophotométrie après coloration de Feulgen (couramment employée jusqu'au début des années 80). Elle donne des estimations de la taille des génomes comparables (Michaelson *et al.*, 1991).

### **6.1. MATERIEL ET METHODE**

#### **6.1.1. Matériel végétal**

Le matériel végétal étudié dans cette partie provient exclusivement de la "core collection" de Montpellier. En effet, pour des raisons techniques (il est extrêmement difficile d'obtenir des résultats convenables en cytométrie de flux lorsque les échantillons sont abîmés) nous avons choisi d'utiliser du matériel frais. Les échantillons analysés appartiennent aux principales espèces des *Coffea* d'origine africaine. Quelques échantillons provenant d'espèces malgaches ont également été analysés, mais le faible nombre de génotypes disponibles n'a pas permis de les joindre aux analyses statistiques. Le matériel analysé est décrit dans le tableau suivant :

ESPECE	Population (origine)	Génotype				
<i>C.arabica</i>	Variété cultivée	Catimor				
	Ethiopienne	ET 1.1	ET 12.2	ET 12.5		
<i>C.brevipes</i>	Mont Cameroun	JA1	JA4	JA6	JA12	JA18
<i>C.canephora</i>	Guinéen	BA3	BA5	BA6		
	Congolais	BB8	BB14	BB19		
	Nana	BC4	BC6			
	Cameroun	BD2	BD7	BD19	BD1	
<i>C.congensis</i>	rep. Centrafricaine	CA4 bis	CA9	CA7		
	Cameroun	CB10	CB22	CB13		
	Congo	CC14 bis	CC22 bis	CC23 bis		
<i>C.eugenoides</i>	Kenya	A 16.1	A 16.2	5	15	
	vitroplants	Eug 1				
<i>C.farafaganensis</i>	Madagascar	Far 1				
<i>C.humilis</i>	Guinée	G2	G13	G16		
	vitroplant	Hum 1				
<i>C.liberica</i>	<i>liberica</i> (Guinéen)	EA3	EA12	EA19		
	<i>dewrevrei</i> (Congolais)	EB3	EB9	EB17		
	Koto (Cameroun)	EC4	EC9	EC11	EC14	EC16
<i>C.millotii</i>	Madagascar	CM1				
<i>C.pseudozanguebariae</i>	Kenya	H4	H5	H7		
<i>C.racemosa</i>	intro. Brésil	IA4	IA9	IA11		
	intro. Tanzanie	IB13				
<i>C.salvatrix</i>	intro. Brésil	LA1				
	intro. Tanzanie	LB1	LB13			
<i>C.sessiliflora</i>	Shimba (Kenya)	PA3	PA4	PA7		
	Kitulangalo (Tanzanie)	PB7	PB10	PB13		
<i>C.costatifructa</i>	Tanzanie	OH1	OH10	OH17	OH20	
<i>C.sp.Moloundou</i>	Cameroun	7bis				
	Souanké (Congo)	1bis	2bis			
<i>C.stenophylla</i>	Ouest CI	FA4	FA6	FA11		
	Est CI	FB3	FB12			

**Tableau 15.** Matériel végétal utilisé pour la détermination de la taille des génomes (en provenance de la "core collection" ORSTOM).

### 6.1.2. Méthode

La technique de cytofluorométrie en flux doit être adaptée à chaque type de matériel végétal. L'étude directe de protoplastes en CMF est envisageable (taille et symétrie convenant au système de "flux") mais il est cependant nécessaire de perméabiliser au préalable la membrane cytoplasmique pour rendre possible la coloration par des fluorochromes exogènes. Cette perméabilisation diminuerait la précision de la mesure de CMF (Dolezel, 1991). Dans la

pratique, ce sont le plus souvent des noyaux isolés qui sont utilisés. Ils sont obtenus soit par lyse des protoplastes, soit par lacération de feuilles dans un tampon lysant (Galbraith *et al.*, 1983; Dolezel *et al.*, 1989; Arumuganathan & Earle, 1991).

Nous avons utilisé deux approches, chacune basée sur l'utilisation d'un fluorochrome particulier ayant des spécificités propres (Dolezel *et al.*, 1992) : L'iodure de propidium (IP) et le 4',6-diamidino-2-phenylindole (DAPI).

L'iodure de propidium (IP) est un intercalant de l'ADN double brin : c'est une grosse molécule qui ne traverse pas la membrane cytoplasmique et nécessite le plus souvent l'emploi d'un détergent pour perméabiliser la membrane nucléaire (Dolezel *et al.*, 1992). Sa propriété de se lier stoechiométriquement à l'ADN en fait le fluorochrome de choix pour les déterminations du contenu en ADN nucléaire. Après plusieurs essais avec les techniques "classiques" d'isolement des noyaux (Dolezel *et al.*, 1989; Dolezel, 1991; Arumuganathan & Earle, 1991), il nous est clairement apparu que les espèces de caféiers avaient des comportements très variables. L'analyse de certaines espèces, voire de certains échantillons, demeurerait impossible (absence de pics ou C.V. très élevés). Nous avons donc adapté un protocole d'isolement des noyaux plus "élaboré" qui a permis d'obtenir des résultats homogènes, quel que soit le génotype étudié (annexe 8).

Dans un deuxième temps nous avons utilisé le 4',6-diamidino-2-phenylindole (DAPI) qui est un fluorochrome spécifique des paires de bases A-T. Sa spécificité rend la valeur observée à la fois dépendante du contenu en ADN mais également de la proportion de bases A-T. Ce type de fluorochrome pénètre plus facilement les membranes et autorise même la coloration de cellules vivantes (non lysées). L'analyse de notre matériel lors de colorations par ce fluorochrome a été possible après un isolement "classique" des noyaux à l'aide d'un tampon lysant (Cros *et al.*, 1994).

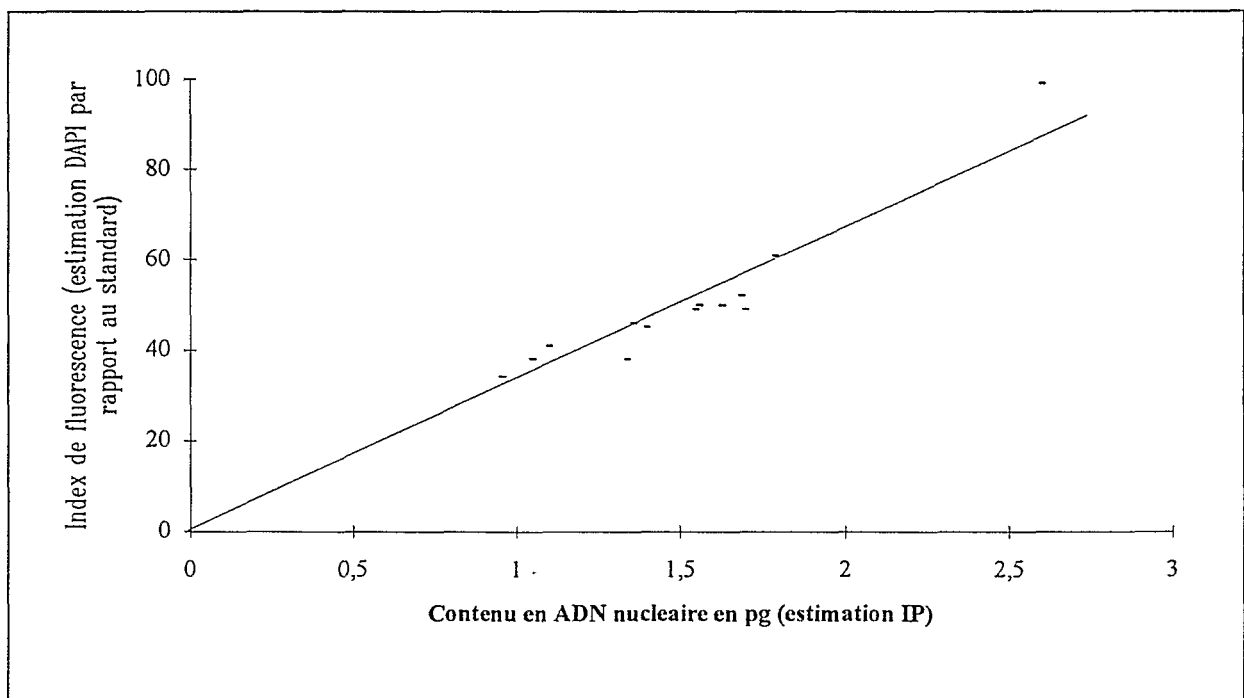
## **6.2. RESULTATS**

L'estimation du contenu en ADN nucléaire des différents échantillons ( $qADN_{ech}$ ) est obtenue à partir des résultats observés au cours des "colorations IP", à l'aide d'un standard dont la

teneur en ADN nucléaire est connue ( $qADN_{std}$ ). La teneur en ADN nucléaire est calculée selon la formule suivante (Arumuganathan & Earle, 1991):

$$qADN_{ech} = (\text{position du pic ech.} / \text{position du pic std.}) \times qADN_{std}$$

Pour la raison précédemment citée, les positions des pics observés au cours des "colorations DAPI" ne sont pas converties en  $qADN$ . Les résultats sont donc donnés en valeurs absolues sur une échelle linéaire, par rapport à un échantillon de *C.arabica* arbitrairement positionné sur 100 (les positions ont été au préalable recalibrées par rapport à un standard interne *Petunia hybrida*). Entre espèces diploïdes, les valeurs arbitraires d'index de fluorescence "DAPI" varient de 34 pour *C.racemosa* à 61 pour *C.humilis*. La figure 22 présente une analyse de régression réalisée sur les valeurs moyennes par espèce. Les estimations réalisées à l'aide de ces deux fluorochromes sont bien corrélées ( $r=0,95$ ). Ce résultat signifie que le pourcentage de paires de bases AT pour ces espèces n'est pas significativement différent. De plus, la droite de régression passe par un point qui n'est pas significativement différent de l'origine (0,0). Il est donc possible, une fois ce travail préalable réalisé, d'estimer le contenu en ADN nucléaire d'un échantillon appartenant à l'une de ces espèces, indirectement à partir de valeurs DAPI (la préparation des échantillons est beaucoup plus rapide).



**Figure 22.** Corrélations des résultats obtenus en cytométrie en flux à partir de mesures réalisées après coloration par le DAPI (sur noyaux non purifiés) et par l'iodure de propidium (sur des noyaux isolés).

### 6.2.1. Taille des génomes

La taille du génome diploïde des caféiers varie de 0,9 à 1,9 pg, ce qui permet de calculer une taille moyenne de 1,64 pg soit environ 1,54 Mpb (pour  $2n=22$  chromosomes). Chez les eucaryotes, la taille des génomes est extrêmement variable : elle est par exemple de 0,33pg pour *Arabidopsis thaliana* (pour  $2n=10$  chromosomes) (Marie & Brown, 1993), de 30,9pg pour le blé (*Triticum aestivum*, hexaploïde) (Marie & Brown, 1993), et chez les angiospermes elle peut atteindre 223pg pour *Trillium rhombifolium* (Liliaceae) (Bennet & Smith, 1991). La taille du génome des caféiers est relativement faible par rapport à la taille du génome d'autres angiospermes, elle apparaît cependant être du même ordre de grandeur que celle d'*Acacia heterophylla* (1.60 pg), de *Beta vulgaris* (1.75 pg), ou de l'igname, *Dioscorea alata* (1.47 pg) (Marie & Brown, 1993).

### 6.2.2. Discriminations de plusieurs groupes de taille de génome chez les caféiers

Les tailles déterminées sont génotypes spécifiques. En effet, d'importantes variations entre les tailles de génomes nucléaires de caféiers appartenant à la même espèce sont enregistrées. Ces variations intraspécifiques semblent toutefois moins importantes que la variation interspécifique de l'ensemble du genre. Afin de tenter de mettre en évidence une structuration de la taille des génomes, une première analyse discriminante a été réalisée sur les valeurs individuelles (par génotype). Cette analyse (utilisant le critère de la variance intragroupe minimale) permet de distinguer 3 groupes de taille de génome (les résultats sont donnés dans le tableau 16). Certaines espèces sont nettement spécifiques d'un groupe alors que d'autres (dont les génotypes ont des tailles beaucoup plus variables) sont présentes dans 2 groupes.

Espèce	fréquence des génotypes dans le groupe		
	0.90 - 1.30pg ADN/ny	1.31 - 1.60pg ADN/ny	1.61 - 1.80pg ADN/ny
<i>C. humilis</i>		0.2	0.8
<i>C. sp. Moloundou</i>		0.25	0.75
<i>C. liberica</i>		0.34	0.66
<i>C. congensis</i>		0.38	0.62
<i>C. brevipes</i>		0.63	0.37
<i>C. canephora</i>		0.64	0.36
<i>C. eugenioides</i>		1	
<i>C. stenophylla</i>		1	
<i>C. costatifructa</i>		1	
<i>C. pseudozanguebariae</i>	0.8	0.2	
<i>C. sessiliflora</i>	1		
<i>C. racemosa</i>	1		

**Tableau 16.** Fréquence des génotypes de chaque espèce dans chacun des 3 groupes d'aggrégation déterminés par le critère de la variance intragroupe minimale.

Pour évaluer la variation intraspécifique par rapport à la variation interspécifique, un test de Newman et Keuls a été réalisé sur les valeurs moyennes par espèce (moyennes réalisées après tirage aléatoire, sur un effectif par espèce constant). Ce test, dont les résultats sont donnés dans le tableau 17, montre qu'il est impossible de discriminer certaines espèces entre elles.

Ce test permet de discriminer trois groupes d'aggrégations. Un premier groupe rassemble des espèces ayant des génomes de taille moyenne mais plutôt élevée (a et a,b), un deuxième groupe rassemble des espèces ayant des génomes de taille moyenne (b) et enfin un troisième groupe, clairement séparé des autres réunit trois espèces à petit génome (c).

Espèce	pléidie	Moyenne par espèce (pg ADN/ny +/- C.V.)	Groupe d'agrégation
<i>C. arabica</i>	4x	2,61 +/- 0,23	(non inclus dans l'analyse)
<i>C. humilis</i>	2x	1,78 +/- 0,33	a
<i>C. sp. Moloundou</i>	2x	1,69 +/- 0,25	a,b
<i>C. liberica</i>	2x	1,68 +/- 0,29	a,b
<i>C. congensis</i>	2x	1,62 +/- 0,19	b
<i>C. brevipes</i>	2x	1,55 +/- 0,18	b
<i>C. canephora</i>	2x	1,54 +/- 0,22	b
<i>C. eugenoides</i>	2x	1,39 +/- 0,12	b
<i>C. stenophylla</i>	2x	1,35 +/- 0,12	b
<i>C. costatifructa</i>	2x	1,33 +/- 0,02	b
<i>C. pseudozanguebariae</i>	2x	1,09 +/- 0,13	c
<i>C. sessiliflora</i>	2x	1,04 +/- 0,16	c
<i>C. racemosa</i>	2x	0,95 +/- 0,13	c

**Tableau 17.** Détermination des valeurs moyennes de la taille des génomes (en pg d'ADN/noyau) pour 13 espèces du sous-genre *Coffea*. Les groupes d'agrégation sont déterminés par un test de Newman et Keuls, les lettres a, b, c indiquent l'appartenance à un groupe de valeurs non significativement différente à  $P=0,05$ .

L'ensemble des deux analyses (par génotypes et par espèces) ne permet de distinguer clairement que 2 groupes.

-Le premier (groupe c et groupe dont les génotypes contiennent entre 0,9 et 1,3pg d'ADN/ny) est très uniforme et regroupe les génotypes des trois espèces, *C. pseudozanguebariae*, *C. sessiliflora* et *C. racemosa*.

-Le second réunit des espèces dont les valeurs moyennes du contenu en ADN sont proches. Ce "groupe" est moins homogène. Trois espèces (*C. eugenoides*, *C. stenophylla* et *C. costatifructa*) ont les valeurs moyennes les plus faibles du groupe b, et la totalité des génotypes étudiés pour ces espèces contiennent entre 1,3 et 1,6 pg ADN/ny. A l'opposé, l'espèce *C. humilis* est bien discriminée dans les deux types d'analyses, la valeur moyenne la place parmi les "gros" génomes de caféiers. Il reste enfin 5 espèces dont les tailles de génome ne peuvent être discriminées. Pour ces 5 espèces, les valeurs moyennes par espèce de la qADN sont très proches (1,54 à 1,69 pg) et la variation intraspécifique est importante.

### **6.3. DISCUSSION**

#### **6.3.1. Variation de la taille des génomes**

Avant le début de cette étude, la seule valeur du contenu en ADN nucléaire connue pour les caféiers concernait un échantillon de *C.canephora* pour lequel on avait déterminé 1,67pg d'ADN/noyau (Marie & Brown, 1993). Cette valeur (bien que cohérente avec les résultats que nous avons obtenus) cache des disparités importantes entre génotypes, appartenant ou non à des espèces différentes.

A l'intérieur même de *C.canephora*, la teneur en ADN nucléaire varie de 1,54 à 1,70pg. Chez certaines espèces de caféiers, la variation intraspécifique atteint donc 15%. Cette variation intraspécifique peut être largement plus importante dans d'autres espèces. Elle atteint 30% chez *Zea mays*, 54% chez *Glycine max*, et jusqu'à 228% chez *Collinsia verna* (Bennet & Smith, 1991). Chez *Helianthus annuus*, une variation supérieure à 58% est couramment observée entre lignées à l'intérieur d'une même variété (Natali *et al.*, 1993). Ces variations seraient dues non pas à un changement du nombre de chromosomes ou à des altérations chromosomiques importantes, mais plutôt à des variations de séquences répétées (Natali *et al.*, 1993). Chez les caféiers, les variations intraspécifiques mesurées sont relativement faibles par rapport à la variation globale au niveau du genre.

Entre l'ensemble des échantillons étudiés, la quantité d'ADN nucléaire varie du simple au double. Ce résultat est en accord avec les observations de Price (1988) qui a noté qu'une variation dans un rapport de 2 à 3 fois entre espèces congénériques était assez courante. Dans le genre *Helianthus*, le contenu en ADN nucléaire d'espèces diploïdes peut même varier d'un facteur 4 (Sim & Price, 1985; Michaelson *et al.*, 1991). Il existe une variation du même ordre entre 8 espèces congénériques du genre *Dioscorea* (igname) mais elle est essentiellement liée à la variation du niveau de ploïdie (Hamon *et al.*, 1992).

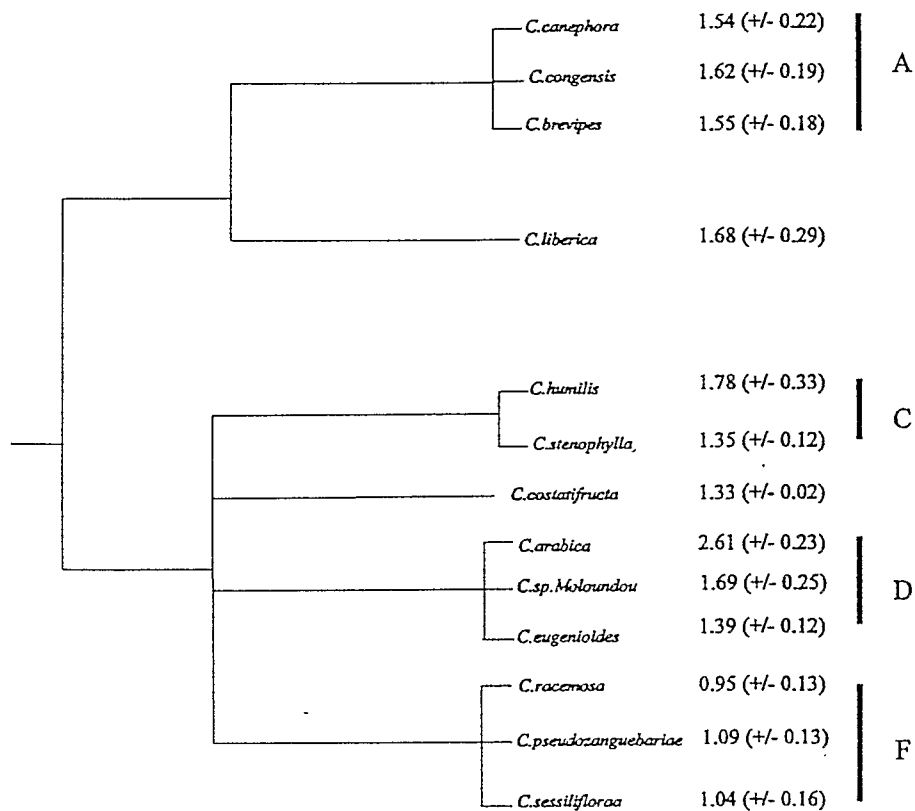
#### **6.3.2. Relation entre variation du contenu en ADN nucléaire et phylogénie**

La variation intraspécifique du contenu en ADN nucléaire observée chez les caféiers est importante. Ce caractère n'est donc à priori pas un bon "marqueur" pour la phylogénie des espèces. Les valeurs moyennes par espèce ont été notées sur l'arbre phylogénétique établi à partir des données chloroplastiques (figure 23). Deux situations s'opposent. Pour les clades A et F déterminés par l'analyse phylogénétique, le contenu en ADN nucléaire est homogène à



l'intérieur du clade. C'est à dire que tout ce passe comme s'il y avait une "tendance" pour l'évolution de la taille des génomes sur un phylum. Au contraire, pour le clade C (et de façon moins nette pour le clade D), les valeurs moyennes (par espèce diploïde) du contenu en ADN nucléaire sont clairement discriminées (appartenance à 2 groupes d'aggrégation différents dans le test de Newman et Keuls). Dans ces cas là, il faut admettre qu'il existe un phénomène supplémentaire qui se "superpose" à la variation de la taille des génomes liée à l'évolution.

Certains auteurs ont suggéré que l'évolution (à un niveau de ploidie donné) soit liée à une augmentation graduelle de la taille du génome (Gupta & Tsuchia, 1991). Dans le genre *Gibasis*, Kenton remarque que 2 espèces morphologiquement et écologiquement très proches ont des contenus en ADN très variables (Kenton, 1984). Cette variation n'est pas interprétable comme un processus d'adaptation aux conditions du milieu, Kenton propose donc une origine liée à l'évolution des espèces. Les espèces du groupe *G.linearis* qui ont des contenus en ADN sensiblement plus élevés sont également les espèces les plus spécialisées. Chez les caféiers, le centre d'origine supposé est l'Afrique de l'est près de la dorsale du Kivu (Charrier, 1976; Leroy, 1982). Globalement, il pourrait donc exister une modification de la taille des génomes depuis le centre d'origine vers les zones "colonisées" au cours de l'évolution (augmentation vers la zone guinéo-congolaise et Madagascar, diminution vers la région centre et sud-est africaine). A ce phénomène se superposerait un phénomène probablement plus rapide de variation de la taille des génomes comme adaptation à l'environnement et aux conditions du milieu.



**Figure 23.** Relations phylogénétiques établies à partir de données chloroplastiques, entre quelques espèces de caféiers dont la taille moyenne du génome nucléaire a été estimée.

### 6.3.3. Relation avec les conditions écologiques

Price (1988) a noté que, chez les angiospermes, la diminution du contenu en ADN nucléaire pourrait être un processus d'adaptation d'un groupe de plantes soumises à des conditions de sécheresse. Cet auteur rapporte également qu'il existe une corrélation positive entre un faible contenu en ADN nucléaire et un cycle cellulaire de courte durée. *C. pseudozanguebariae*,

*C. sessiliflora* et *C. racemosa* sont trois espèces qui présentent une morphologie et un comportement assez proche : elles portent un faible nombre de fleurs par noeud, ont un cycle de maturation des fruits de 2 mois environ (au lieu de 7 à 15 mois pour les espèces guinéo-congolaises), et portent de petits fruits (Hamon *et al.*, 1984; Anthony, 1992). Chez 19 espèces de pins nord américains, on a noté une corrélation positive entre la taille du génome nucléaire, l'indice de croissance et la taille des graines (Wakamiya *et al.*, 1993). Plus récemment, Graham *et al.* (1994) ont montré que chez le soja (*Glycine max*) il existait une corrélation significative entre la taille du génome (au même niveau de ploïdie) et le groupe de maturité. Par ailleurs, il semble exister une forte corrélation négative avec une faible moyenne de précipitation annuelle (Wakamiya *et al.*, 1993). Cette interprétation pourrait expliquer la différence entre les valeurs moyennes de la teneur en ADN par noyau des deux espèces du clade C (*C.stenophylla* et *C.humilis*). *C.stenophylla* occupe une position charnière. Bien qu'elle possède des caractères phénologiques communs avec les autres espèces guinéo-congolaises, cette espèce est adaptée à de relatives conditions de sécheresse des sommets de collines de sa zone d'endémisme, à l'extrême bordure ouest de la Guinée et du nord de la Côte d'Ivoire (Berthaud, 1984). Au contraire, l'espèce *C.humilis* est adaptée à des conditions climatiques plus humides (Berthaud, 1984). Bien que moins contrastée, on retrouve une situation équivalente entre les 2 espèces diploïdes du clade D. *C.sp.Moloundou* est un taxon originaire de la forêt tropicale de plaine (altitude 400m environ) du sud Cameroun - nord Congo. *C.eugenioides*, au contraire, est un taxon afro-montagnard (alt. 1600 à 2000m). L'adaptation de *C.sp.Moloundou* à des conditions plus humides que *C.eugenioides* semble ici encore être corrélée à une teneur en ADN nucléaire plus importante.

Chez les caféiers, il semble donc qu'une taille de génome plus faible que la "moyenne" puisse faire partie d'un ensemble de phénomènes d'adaptation à un milieu plus sec. Des espèces phylogénétiquement proches pourraient ainsi avoir des tailles de génomes très différentes.

#### **6.4. BILAN**

Les caféiers ont une taille de génome diploïde comprise entre 0,9 et 1,9pg d'ADN par noyau. Cette taille les place plutôt parmi les Angiospermes à petit génome.

Les variations du contenu en ADN nucléaire à l'intérieur du genre (au niveau diploïde) vont du simple au double (100%), et les variations intraspécifiques peuvent atteindre 20%. La variance intraspécifique de certaines espèces est telle que toutes les espèces ne sont pas discriminées.

Les variations des tailles des génomes nucléaires pourraient être liées à 2 phénomènes parallèles : d'une part l'évolution sur un phylum donné (la teneur en ADN nucléaire de toutes les espèces d'un même clade évoluerait dans le même sens), et d'autre part un phénomène plus "rapide" d'adaptation à des conditions écologiques particulières (par exemple, diminution de la teneur en ADN pour les espèces adaptées à des conditions de sécheresse).

## **7.DISCUSSION CONCLUSION**

Le sous genre *Coffea*, qui comporte environ 80 taxons, est le plus important de la tribu des *Coffeae*. Son découpage en 3 sections "*Erythrocoffea*" "*Mozambicoffea*" et "*Mascarocoffea*" a été proposé principalement sur la base de la répartition géographique des espèces (Chevalier, 1947). Des études ultérieures concernant les aspects cytogénétiques ou la diversité biochimiques ont permis de proposer un découpage qui reprend globalement les limites des séries initialement proposées par Chevalier (Berthaud & Charrier, 1988; Anthony *et al.*, 1993).

L'objectif de notre travail était de préciser l'organisation évolutive et les relations de parenté entre espèces de caféiers (dont principalement celles appartenant au sous-genre *Coffea*) c'est-à-dire de proposer une phylogénie des espèces de ce genre en relation avec l'histoire du continent africain.

L'approche choisie a été l'étude du polymorphisme du génome chloroplastique, complétée de données concernant les variations de taille du génome nucléaire. Les études moléculaires fournissent en effet de nouveaux marqueurs génétiques qui produisent des phylogénies généralement bien résolues (Weir *et al.*, 1990). L'ADNcp produit, de plus, des reconstructions phylogénétiques comportant peu d'homoplasie (Clegg & Zurawski, 1992; Olmstead & Palmer, 1994; Hoot *et al.*, 1994; Van Ham & Hart, 1994).

Nous rappellerons brièvement les résultats obtenus, sans revenir sur les discussions abordées à la fin de chacun des chapitres précédents. Nous développerons ensuite une discussion autour des informations nouvelles qu'apporte la synthèse de ces résultats quant à l'origine (localisation et époque) et la spéciation des caféiers.

## **7.1. SYNTHÈSE DES RESULTATS**

### **Taille et variation de taille des génomes chloroplastique et nucléaire**

Les résultats obtenus concernent en premier lieu la détermination des tailles des génomes chloroplastiques et nucléaires de plusieurs espèces de caféiers.

La taille déterminée pour l'ADNcp est d'environ 180 kb. C'est une taille moyenne chez les plantes supérieures mais plutôt élevée par rapport aux angiospermes (Palmer, 1985; Sugiura, 1992) et particulièrement aux autres genres de Rubiaceae (Bremer & Jansen, 1991). Cette taille paraît très peu variable entre toutes les espèces étudiées.

Une taille de génome nucléaire comprise entre 0,9 et 2,7 pg d'ADN/noyau est observée chez les caféiers (Cros *et al.*, 1994). Cette taille est plutôt faible par rapport aux tailles des génomes nucléaires observées chez les angiospermes (Marie & Brown, 1993). Chez les caféiers diploïdes, cette taille varie du simple au double (0,9 à 1,8 pg). Cette différence pourrait traduire une adaptation à des biotopes différents (Kenton, 1984; Wakamiya *et al.*, 1993). Ce phénomène pourrait intervenir dans les mécanismes de spéciation comme cela a été montré pour les genres *Gibasis* (Kenton, 1984), *Vaccinium* (Costich *et al.*, 1993) *Lolium*, *Lathyrus* et *Bulnesia* (voir Gupta & Tsuchia, 1991 pour une revue).

### **Divergence intragénérique**

Le niveau de variation de l'ADN chloroplastique des caféiers évalué par l'analyse du polymorphisme des sites de restriction de l'ensemble du génome chloroplastique est inférieur à 1%. La divergence de la séquence chloroplastique non codante trnL-trnF ne semble pas, quant à elle, dépasser 3%. Le polymorphisme intragénérique apparaît donc faible chez les caféiers par rapport aux résultats obtenus pour d'autres genres (Palmer *et al.*, 1988; Dowling *et al.*, 1990; Hillis *et al.*, 1990). Cette divergence réduite se traduit par un faible nombre de caractères phylogénétiquement informatifs en regard du nombre de caractères étudiés (tous les génotypes ne sont pas discriminés).

### **Analyse phylogénétique**

L'interprétation phylogénétique des résultats produit un arbre ayant une topologie "en râteau", signe d'une explosion radiative rapide. L'étude de 52 génotypes appartenant à 27 taxons de caféiers a permis de distinguer au moins 6 clades réunissant des espèces contemporaines possédant un génome chloroplastique ancêtre commun. Les espèces présentes dans chaque clade ont des aires de distribution qui se superposent et d'étroites relations de parenté. Cette étude permet également de positionner 6 taxons indéterminés (*C.sp.*). L'observation d'un phénomène d'hétérogénéité intragénotype de l'ADNcp (pour certains caractères et certains individus) pourrait suggérer une ascendance commune entre caféiers endémiques de part et d'autre du canal du Mozambique.

## 7.2. DISCUSSION GENERALE

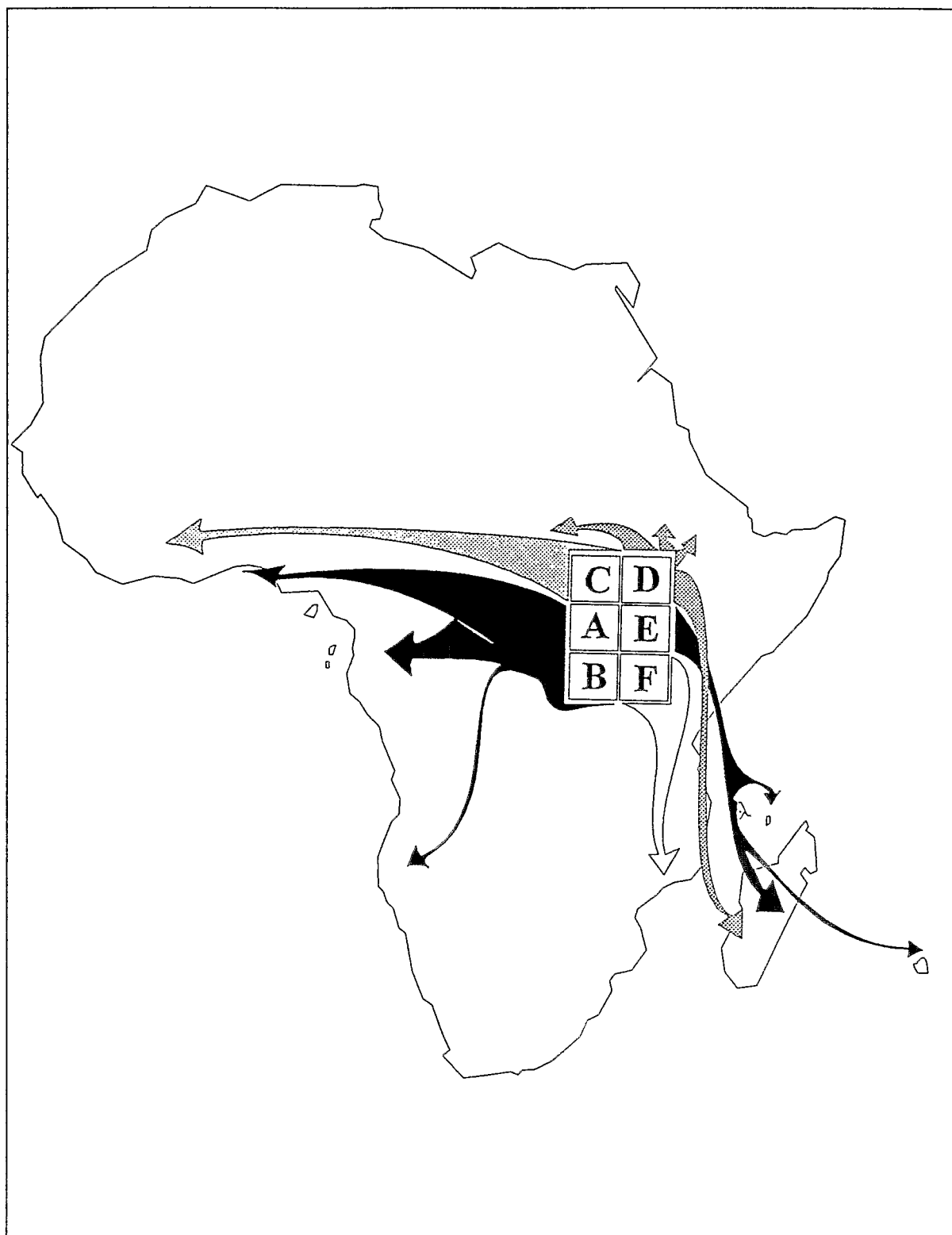
### Centre d'origine

La notion de centre d'origine (ou de diversification) est définie par Vavilov comme étant la zone géographique concentrant le maximum de diversité (morphologique, cytologique, génétique...). Pour l'Afrique il a identifié deux centres d'origine des plantes cultivées, l'Ethiopie et le pourtour méditerranéen (Chauvet & Olivier, 1993). Le plateau abyssin, qui va de l'Ethiopie au Kenya, en bordure nord ouest de la vallée du rift, est connu pour être un important centre de diversification pour de nombreuses espèces végétales sauvages et cultivées (Harlan, 1969). Cette zone a en effet subi des variations climatiques propices à la spéciation. Un important système de fractures accompagné de volcanisme s'est mis en place au cours du tertiaire. Il a produit ce que l'on désigne sous le nom de "grands rifts africains" qui dessinent aujourd'hui 2 fractures principales: une grande coupure nord-sud qui va de la mer Rouge (dépression de l'Arar) au Kilimandjaro, constitue la "vallée du rift" ; une branche occidentale, tectoniquement plus active, passe par les grands lacs (Moukoko Sese Seko, Kivu, Tanganyika et Malawi), et constitue la "dorsale du Kivu". La zone du lac Victoria est comprise entre ces deux fractures. Au cours du quaternaire cette zone a subi des variations climatiques importantes (en relation avec les variations similaires de l'Europe) et a été la région d'apparition de l'homme.

Chez les caféiers, les analyses cytogénétiques ont montré une différenciation chromosomique peu importante entre espèces africaines et malgaches. Ce résultat a permis de proposer l'hypothèse d'une origine commune des caféiers (Charrier, 1976). Sur la base des mêmes résultats, un centre de "diversification du génome" situé dans la zone des rifts est-africains a été proposé (Demarly, 1975). Un argument supplémentaire était que, lors d'hybridations interspécifiques, l'espèce *C.eugenioides*, endémique de cette zone, occupait une position charnière : elle semble en effet s'hybrider aussi bien avec certaines espèces malgaches, guinéo-congolaises ou provenant d'Afrique de l'est (Charrier, 1976).

Les aires de trois des six clades déterminés se superposent dans la région centrale de l'Afrique. Sur la base de ces données, on peut proposer la zone comprise entre les 2 dorsales (vallée du rift et dorsale du Kivu) comme étant la zone géographique concentrant le maximum de diversité en terme de chlorotypes. Nos résultats concernant les variations de l'ADNcp renforcent donc l'hypothèse d'un centre d'origine des caféiers dans cette zone. La figure 24 schématise les voies possibles de diversification à partir de cette zone d'origine supposée.





**Figure 24.** Voies proposées pour la diversification à partir d'un centre d'origine situé entre les 2 zones de fractures d'Afrique de l'est. Les lettres (A,B,C,D,E,F) font référence aux clades définis sur la figure 18.

## Spéciation des caféiers

La phylogénie construite à partir des données chloroplastiques suggère l'existence d'au moins 6 clades entre les 27 taxons contemporains étudiés.

- Les deux premiers clades (A,B) correspondent respectivement aux canéphoroïdes (*C.canephora*, *C.congensis*) auxquels sont associés l'espèce *C.brevipes* et le taxon camerounais *C.sp.N'koumbala*, d'une part (clade A) ; et aux Liberio-excelsoïdes auxquels sont associés deux taxons congolais endémiques de la forêt du Mayombe (*C.sp.* Ngongo II et *C.sp.* Mayombe), ainsi que les taxons *C.sp.X* et *C.kapakata*, d'autre part (clade B). Ces deux premiers clades regroupent des espèces couvrant largement la zone guinéo-congolaise.

- Le troisième clade (C) aboutit aux deux espèces *C.stenophylla* et *C.humilis*, aujourd'hui strictement endémiques de la zone guinéenne. Au cours des prospections récentes en Afrique centrale, aucune forme congolaise de ces espèces n'a été retrouvée. Ce clade qui apparaît fortement supporté par de nombreuses synapomorphies, pourrait provenir d'une migration de l'espèce ancêtre à l'époque du Miocène. Ces deux espèces sont phénotypiquement très différentes. La taille de leur génome est également très différente bien qu'elles aient le même niveau de ploïdie. La spéciation de ces deux espèces à partir de leur ancêtre commun a très certainement eu lieu récemment (par rapport à la spéciation des caféiers) en s'accompagnant d'une évolution différentielle de la taille des génomes. Nous avons vu (paragraphe 6.3.3.) que la variation de la taille du génome pourrait être corrélée, dans plusieurs cas, avec une adaptation à des conditions climatiques et d'environnement. Une augmentation de la taille du génome nucléaire prévaudrait dans les zones les plus humides ou, au contraire, une diminution de la taille existerait dans les zones plus sèches.

- Le quatrième clade (D) rassemble les espèces contemporaines *C.eugenioides* et *C.arabica*, originaires de la région de la dorsale du Kivu et *C.sp.Moloundou* récemment collecté dans 5 populations à la frontière Cameroun - Congo (Anthony, 1992). Bien qu'isolé, *C.sp.Moloundou* possède des caractéristiques qui rappellent celles des taxons est-africains (forme arbustive, petites feuilles, petits grains faiblement chargés en caféine); de plus ce taxon est auto-compatible comme *C.arabica*. La présence de *C.sp.Moloundou* dans le clade D est le signe qu'il n'est phylogénétiquement pas lié aux taxons sud orientaux du clade F dont il est morphologiquement proche. L'histoire de *C.sp.Moloundou* serait au contraire étroitement liée à celle de *C.eugenioides* et de *C.arabica*. La différenciation de ces trois espèces serait un phénomène récent qui pourrait avoir eu lieu au cours du quaternaire.

- Un cinquième clade (F) associe l'histoire de 3 taxons planitaires des zones côtières est-africaines (*C.racemosa*, *C.pseudozanguebariae*, *C.sessiliflora*) et un taxon spontané à la frontière Mozambique - Zimbabwe, à une altitude plus élevée (*C.salvatrix*). Une autre voie de spéciation pourrait être préfigurée par *C.costatifructa*, espèce endémique d'une île côtière tanzanienne, qui occupe une place à part. Il ne partage pas les synapomorphies du clade regroupant les quatre espèces décrites ci-dessus, et il ne possède qu'une seule autapomorphie (voir figure 18). Ces deux clades (F + *C.costatifructa*) regroupent des espèces présentes dans la zone est-africaine. La spéciation de trois espèces (au moins) de cette zone (*C.racemosa*, *C.pseudozanguebariae*, *C.sessiliflora*) est associée à une diminution du contenu en ADN nucléaire, sans que l'on soit capable de préciser, encore une fois, si cette diminution de la quantité d'ADN nucléaire en est la cause ou la conséquence.

La diversité morphologique des caféiers dans cette région est importante mais comparable à celle relevée pour d'autres espèces végétales. Elle serait liée à la répartition de la végétation selon des niveaux d'altitude différents. Le relief (nombreux massifs montagneux d'altitude élevée) et le climat (régime de pluies locales et intermittentes) de cette zone auraient, en effet, provoqué des pressions de sélection diversificatrices. L'ensemble de ces phénomènes a conduit les botanistes à décrire un très grand nombre de taxons : *C.kimbozensis*, *C.lulandoensis*, *C.schliebenii*, *C.pocsii* et *C.costatifructa* (Bridson, 1994), correspondant respectivement aux *C.sp.A*, *C.sp.C*, *C.sp.D*, *C.sp.E*, *C.sp.F* précédemment présentés (Bridson & Verdcourt, 1988). De plus, certains de ces taxons n'ont pas encore de statut spécifique (*C.sp.B*, *C.sp.G*, *C.sp.H* et *C.sp.I*). L'étude des relations de parenté entre caféiers ouest et est africains mériterait d'être confirmée par l'étude de taxons tels que *C.fadenii*, *C.mongensis* et *C.mufindiensis*, mais également par l'étude des espèces récemment décrites.

-Le sixième clade (E) aurait donné la plupart des espèces contemporaines à Madagascar comme : *C.farafaganensis*, *C.bertrandii*, *C.millotii* et *C.pervilleana*. L'espèce *C.humblotiana*, endémique des Comores, serait également issue de ce groupe.

L'échantillonnage de *Mascarocoffea* que nous avons étudié, bien que représentatif, est limité par rapport aux quelques 50 taxons de *Mascarocoffea* décrits. L'importante diversité morphologique de ces espèces serait liée à leur diversification par cladogénèse, faite d'isolements successifs, comme cela a déjà été suggéré par Charrier (1978). Les *Mascarocoffea* sont endémiques à Madagascar (île séparée de l'Afrique depuis environ 80 MaBP), mais sont également représentés dans l'archipel des Comores (par une espèce, *C.humblotiana*) et dans les îles Mascareignes par au moins 3 espèces, *C.mauritiana*, *C.macrocarpa*, *C.myrtifolia* (Bossier *et al.*, 1989). Ces îles volcaniques datent du quaternaire (2 à 4 MaBP) et ont donc été colonisées malgré une séparation de près de mille kilomètres. La colonisation d'îles récentes rend caduque l'hypothèse selon laquelle les *Mascarocoffea* auraient nécessairement préexisté à la dislocation gondwanienne (Leroy, 1982). La colonisation de Madagascar pourrait être plus

récente que ce qui avait été suggéré. La diversité morphologique importante qui y est observée aurait été acquise rapidement à l'échelle de la spéciation des caféiers.

Indépendamment de la diversité morphologique, et malgré un échantillonnage réduit, nos résultats montrent que les espèces présentes dans la zone malgache proviennent d'au moins deux voies de spéciation distinctes. En effet, l'histoire ancienne de l'espèce *C.perrieri* apparaît associée à celle du clade D, mais elle aurait divergé assez tôt. L'étude des compositions biochimiques des *Mascarocoffea* avait permis de décrire un "chimiotype perrieri" (Rakotomalala *et al.*, 1993), biochimiquement isolé et très évolué (présence de composés ultimes de la chaîne de biosynthèse des acides cinnamiques). La différenciation de ce taxon pourrait donc être le signe de son histoire indépendante de celle des autres taxons malgaches étudiés.

L'ensemble des clades décrits précédemment est cohérent avec la répartition actuelle des taxons. Les espèces issues des trois premiers (A,B,C) sont strictement localisées dans la zone guinéo-congolaise. Celles issues du clade E ou F sont strictement endémiques de Madagascar et de l'Afrique de l'est, respectivement. Seul le clade D réunit des espèces dont la répartition ne correspond pas à une seule zone. Les taxons présents dans les zones guinéenne et congolaise ont donc, malgré l'homogénéité qui leur avait été attribuée sur la base de caractères morphologiques, une histoire très différente. A l'opposé, et malgré leur apparente diversité morphologique, les taxons orientaux semblent avoir une histoire commune.

### Histoire évolutive

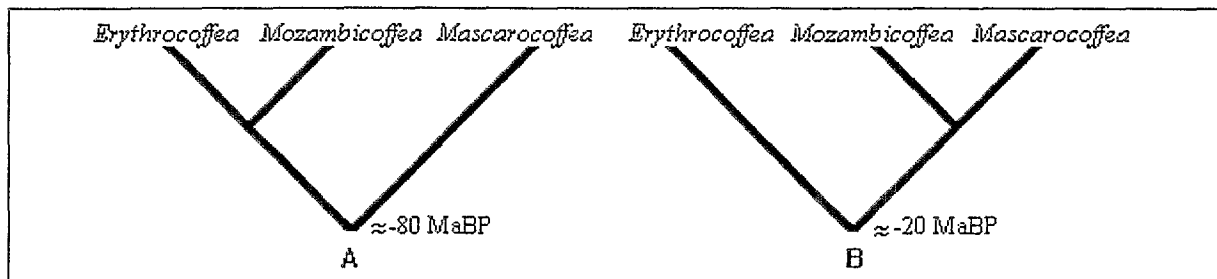
Les résultats concernant la diversité génétique et sa structuration en 3 groupes avaient été analysés comme le signe d'une différenciation primitive en 2 lignées, suivie d'une différenciation de l'une des deux "branches" (voir figure 25) :

-Selon Leroy (1982) la première différenciation aurait eu lieu dans la zone du lac Victoria. Elle consisterait en un isolement primitif des *Mascarocoffea* à Madagascar au moment de la dislocation gondwanienne. Il aurait été suivi par la différenciation d'une lignée vers l'Afrique de l'ouest (libério-excelsoïdes et plus généralement *Erythrocoffea*) et une autre vers l'Afrique de l'est (*Mozambicoffea*). La première différenciation aurait eu lieu à la fin du Crétacé (environ 80 MaBP).

-Au contraire, selon Berthaud (1984), la première différenciation aurait eu lieu à l'intérieur du Kenya et aurait donné une lignée guinéo-congolaise (*Erythrocoffea*), d'une part, et mascaro-mozambicaine, d'autre part. La lignée mascaro-mozambicaine se serait elle même

différenciée en une branche strictement malgache (*Mascarocoffea*) et une autre, tanzano-mozambicaine (*Mozambicoffea*). Ce second modèle suggère une première différenciation aux environs du Miocène (entre 5 et 25 MaBP), c'est à dire au cours de la formation des grands rifts (Berthaud, 1984).

Dans ces deux modèles, les "échanges" plus récents entre ces trois groupes sont introduits pour expliquer les affinités phénotypiques ou génétiques entre espèces appartenant à des groupes de diversité différents.

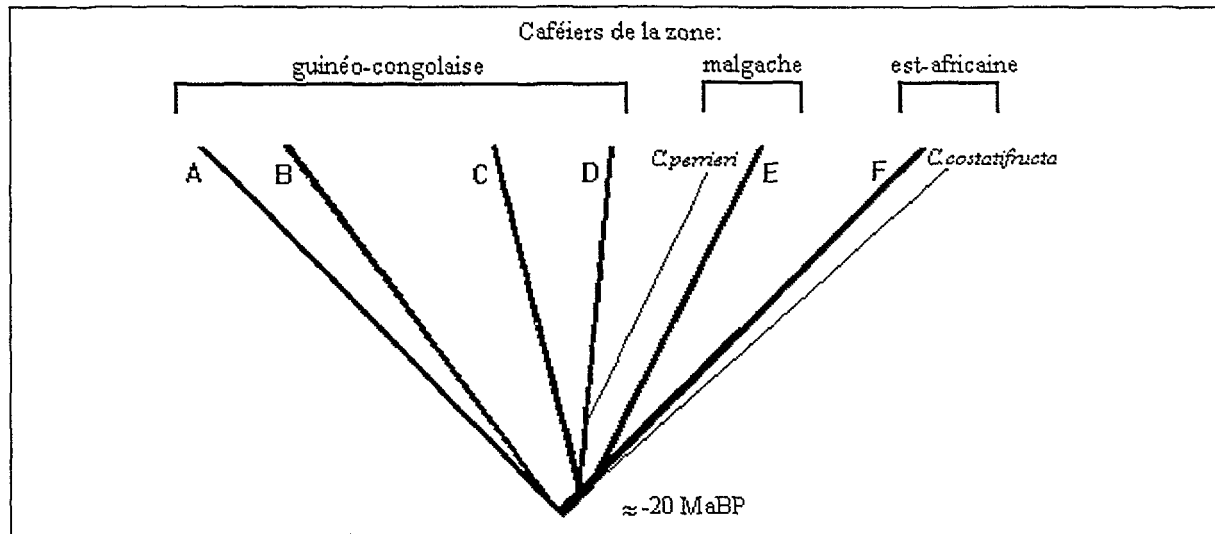


**Figure 25 .** Histoire évolutive des caféiers telle qu'elle était proposée à partir de données concernant la structuration de la diversité génétique en trois groupes. A d'après Leroy (1982), d'après Berthaud (1984).

Les clades décrits à partir de la phylogénie basée sur les variations de l'ADNcp sont compatibles avec la répartition actuelle des espèces de caféiers, mais ne correspondent pas totalement aux trois groupes biogéographiques précédemment définis (Berthaud & Charrier, 1988; Anthony, 1992). L'histoire évolutive suggérée par nos résultats diffère donc des deux "modèles" décrits ci-dessus. La séparation est-ouest n'est pas aussi stricte que les résultats de biosystématique semblaient l'indiquer. La spéciation des caféiers résulte probablement à la fois de phénomènes anciens (plusieurs Ma) et plus récents (moins de 1 Ma) :

- Le premier phénomène est une explosion radiative. Si l'on accepte le postulat d'horloge moléculaire, soit des taux variant de 5 à 30.10<sup>-10</sup> substitutions/site/an pour les séquences non codantes de l'ADNcp (Kawagoe & Kikuta, 1991; Zurawski & Clegg, 1987), cette explosion radiative aurait eu lieu au cours du Miocène (au maximum 30 MaBP); elle aurait été rapide (entre 0,5 et 4 Ma) et aurait produit au moins 6 clades (figure 26). Ce premier phénomène de spéciation serait donc associé à la formation des grands rifts africains, auxquels il se superpose à la fois dans l'espace et dans le temps.

- Le deuxième phénomène de spéciation serait lié aux variations climatiques importantes au cours du quaternaire, l'histoire des refuges forestiers tropicaux (Livingstone, 1982) du quaternaire intervenant dans la distribution actuelle des espèces.



**Figure 26.** Modèle proposé de l'histoire évolutive des caféiers. A, B, C, D, E et F correspondent aux clades décrits dans le texte.

Dans ce modèle, nous ne faisons pas intervenir d'échanges génétiques pour expliquer les convergences phéno-morphologiques entre espèces appartenant à des clades différents.

### 7.3 PERSPECTIVES

La phylogénie que nous proposons entre espèces de caféiers présente des limites liées, d'une part, à l'approche méthodologique (en relation avec le faible nombre de marqueurs polymorphes mis en évidence) et, d'autre part, à l'échantillonnage des espèces étudiées.

#### **Méthodologie**

Afin de valider la phylogénie proposée il serait nécessaire d'augmenter le nombre de caractères phylogénétiquement informatifs (synapomorphies). Dans notre analyse des variations de l'ADNcp plusieurs approches ont été envisagées, dont l'étude de régions non codantes. Pourtant la divergence intragénérique observée ne dépasse pas 3%. En conséquence, il est probable que les variations que l'on pourra mettre en évidence sur le génome chloroplastique

des caféiers ne seront pas suffisantes pour reconstruire une phylogénie plus robuste que celle proposée ici.

Les variations de l'ADN nucléaire, dont le mode d'évolution est beaucoup moins conservateur que celui de l'ADNcp, mériteraient donc d'être étudiées. Deux premières tentatives de mise en évidence de variations de l'ADN nucléaire dans l'optique d'établir des relations de parenté entre espèces ont déjà été envisagées :

- La première concernait un nombre réduit de taxons pour lesquels on a étudié les variations du génome nucléaire par RAPD<sup>1</sup> (Lashermes *et al.*, 1993). Cette approche donne chez les caféiers de bons résultats pour l'étude de la variation intra-spécifique, et pour le suivi d'introgessions (Lashermes *et al.*, 1993; Orozco-Castillo *et al.*, 1994). L'application de cette technique aux génotypes de caféiers décrits au chapitre 3 produit des profils très variables. Aucune affinité entre espèces ne peut en être déduite. La technique des RAPD, très intéressante comme outil de "fingerprint", se révèle donc inadéquate chez les caféiers pour la mise en évidence de relations de parenté entre espèces de caféiers.

- La seconde approche pour la mise en évidence de variations de l'ADN nucléaire, est une analyse du polymorphisme de longueur de fragments de restriction à l'aide de sondes nucléaires (Lashermes *et al.*, 1994). Les sondes "séquences uniques" proviennent d'une banque génomique de *C.arabica* et sont, par ailleurs, utilisées pour le programme de la cartographie du génome de *C.canephora* (Lashermes, communication personnelle). Les signaux obtenus sont très polymorphes. Les résultats donnent des arguments en faveur d'une origine autotétraploïde de *C.arabica*. Si *C.arabica* est un allotétraploïde, les deux espèces "parentes" sont, de toute façon, très proches. Par contre, l'étude des relations entre espèces se heurte à une variabilité intraspécifique très importante.

Afin d'obtenir des caractères phylogénétiquement informatifs et aisément analysables pour l'étude des relations de parenté entre espèces, il est donc apparu intéressant de s'attacher aux études de séquences nucléaires. Elles passent pour plus variables que les séquences des génomes cytoplasmiques. Les zones codantes et non codantes des unités ribosomiques présentent, par ailleurs, un grand intérêt pour les reconstructions phylogénétiques (Hamby & Zimmer, 1992). Leur polymorphisme est théoriquement supérieur à celui de l'ADNcp et "l'évolution horizontale" les préserve de l'hétérogénéité intraspécifique que l'on rencontre chez les autres gènes nucléaires (Doebley, 1992). Pour obtenir des caractères suffisamment polymorphes c'est le séquençage de régions non codantes qui semble le plus adéquat. Le

---

<sup>1</sup>Random Amplified Polymorphic DNA (amplification par PCR à l'aide de courtes amorces de séquences arbitraires)

séquençage préliminaire de l'ITS1 (espaceur interne transcrit à gauche du gène codant pour le 5.8S) chez quelques génotypes montre une proportion de 14 événements mutationnels entre *C.canephora* et *C.millotii* sur 190 pb (environ 7,4 % de divergence); celui de l'ITS2 (espaceur interne transcrit à droite du gène codant pour le 5.8S), 31 événements entre *C.salvatrix* et *P.ebracteolatus* sur 206 pb (environ 15% de divergence).

Ces premiers résultats montrent un polymorphisme important qui pourrait être suffisant pour assurer une reconstruction robuste de la phylogénie des caféiers. Il serait donc très intéressant de poursuivre ce travail afin d'établir un arbre phylogénétique sur la base des variations de l'ADN ribosomique. Cette phylogénie pourra être confrontée à l'arbre phylogénétique que nous avons établi à partir de données chloroplastiques. Ce type de comparaison (phylogénie à partir de données nucléaires d'une part et cytoplasmique d'autre part) fournit des éléments informatifs pour la reconstruction de l'histoire des espèces.

### Echantillonnage

La deuxième voie à approfondir concerne les nouvelles espèces à étudier.

Dans le sous-genre *Coffea* tout d'abord, certaines espèces décrites n'ont pas été étudiées, soit par choix (les 50 espèces de *Mascarocoffea* mériteraient une étude particulière), soit par indisponibilité du matériel végétal (espèces récemment décrites en Afrique de l'est). Certaines de ces espèces mériteraient d'être étudiées pour répondre aux incertitudes de la phylogénie, c'est le cas par exemple des espèces est-africaines *C.fadenii*, *C.mufindiensis*, *C.mongensis*.

Certaines zones africaines n'ont jamais été prospectées, et la découverte récente de "nouvelles" espèces au Cameroun ou au Congo (comme la très intéressante espèce autogame *C.sp.Moloundou*) et au Kenya ou en Tanzanie (*C.pseudozanguebariae*) laisse présager la richesse probable de ces régions. La découverte de nouvelles espèces ou de formes centro-africaines, et l'étude de leur parenté avec les espèces étudiées ici permettraient de confirmer les voies de spéciations suggérées. Il apparaît donc nécessaire d'entreprendre de nouvelles collectes par exemple en Ouganda, au sud du Soudan, au Zaïre et plus généralement dans la zone de la dorsale du Kivu. Ces collectes devraient pouvoir être initiées dans le cadre du récent projet A.C.R.N./I.P.G.R.I.<sup>2</sup> (Charrier communication personnelle).

---

<sup>2</sup> "African Coffee Research Network" / projet de l'Institut International des Ressources Phytogénétiques.



L'intégration d'espèces représentant tous sous-genres voisin du sous-genre *Coffea* (*Afrocoffea*, *Psilanthus* et *Baracoffea*) permettrait probablement de réviser la systématique de l'ensemble des Rubiacées présentant la placentation cofféenne. Cette analyse pourrait être envisagée, par exemple à partir de l'étude de la séquence intergénique trnL-trnF. Un des principaux obstacles actuels pour développer cet aspect est lié à l'accès au matériel végétal.

#### 7.4. CONCLUSION

Notre travail a permis de proposer la première phylogénie moléculaire des caféiers. La phylogénie reconstruite à partir des données chloroplastique permet de donner une interprétation nouvelle de l'histoire des caféiers. L'introduction à notre étude de taxons peu ou pas décrits permet de proposer des hypothèses intéressantes, dont l'étroite parenté entre *C.arabica* et *C.sp.Moloundou*. Cette étude démontre néanmoins que notre connaissance de l'histoire des caféiers est encore imparfaite. La poursuite des prospections et l'étude des nouveaux taxons collectés se révèlent donc indispensables.

En ce qui concerne la classification des espèces notre approche représente une mise en place de l'outil méthodologique. Elle méritera d'être approfondie et revue avec l'aide de botanistes avant d'avoir des conséquences plus générales sur la systématique des caféiers.

De façon plus générale, ce travail constitue un exemple d'étude de phylogénie intragénérique concernant des espèces pérennes de la forêt tropicale africaine. Il constitue un élément supplémentaire dans la compréhension des flores africaines et malgache ainsi que leurs relations.

## REFERENCES BIBLIOGRAPHIQUES

## Références bibliographiques

- Anonyme. 1991. Dix ans de biotechnologie. *Biofutur*, Janvier 1991 : 33-41.
- Anthony F. 1992. Les ressources génétiques des caféiers. Collecte, gestion d'un conservatoire et évaluation de la diversité génétique. *ORSTOM Editions*, TDM n°81, 320 pp.
- Anthony F., Clifford M. N., Noirot M. 1989. La diversité biochimique dans les genres *Coffea* et *Psilanthus*. *Ass. Sc. Inter. Café*, 13ème Colloque, Paipa, Colombie (21-25 août 1989) : 474-484.
- Anthony F., Clifford M. N., Noirot M. 1993. Biochemical diversity of the genus *Coffea* L. : chlorogenic acids, caffeine and mozambioside contents. *Genet. Res. Crop. Evol.*, 40 : 61-70.
- Appels R., Dvorak J. 1982. The Wheat Ribosomal DNA spacer region: Its structure and variation in populations and among species. *Theor. Appl. Genet.*, 63 : 337-348.
- Arumuganathan K., Earle D. 1991 a. Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter*, 9 : 208-218.
- Arumuganathan K., Earle D. 1991 b. Estimation of nuclear DNA content of plant by flow cytometry. *Plant Molecular Biology Reporter*, 9 : 229-241.
- Baldauf S. L., Palmer J. D. 1990. Evolutionary transfer of the chloroplast *tufA* gene to the nucleus. *Nature*, 344 : 262-265.
- Barriel V. 1994. Alignement de séquences et codage des insertions-délétions de nucléotides. Compte rendu du colloque "Méthodes Mathématiques et Informatiques pour la Reconstruction Phylogénétique", Montpellier (19-20 mai 1994).
- Bennet M. D., Smith J. B. 1991. Nuclear DNA amount in angiosperms. *Phil. Trans. R. Soc. Lond.*, 334 : 309-345.
- Berthaud J. 1984. Les ressources génétiques pour l'amélioration des caféiers africains diploïdes. Evaluation de la richesse génétique des populations sylvestres et de ses mécanismes organisateurs. Conséquences pour l'application, Thèse, Université d'ORSAY, 322 pp.
- Berthaud J., Anthony F., Le Pierrès D. 1984. Les caféiers de la Nana. Résultats des observations faites en collection en Côte d'Ivoire. *Café Cacao Thé*, XXVIII(n°1 janv. 84) : 3-11.
- Berthou F., Mathieu C., Vedel F. 1983. Chloroplast and mitochondrial DNA variation as indicator of phylogenetic relationships in the genus *Coffea* L. *Theor. Appl. Genet.*, 65 : 77-84.
- Berthou F., Trouslot P., Hamon S., Vedel F., Quetier F. 1980. Analyse en électrophorèse du polymorphisme biochimique des caféiers. *Café Cacao Thé*, 24 (4) : 313-325.
- Birky C. W. 1990. Evolution and population genetics of organelle genes: mechanisms and models. In Selander R. K., Clark A. J., Whittam T. S. (eds.) "Evolution at the molecular level", Sunderland. pp. 112-134.
- Bosser J., Cadet T., Guého J., Marais W. 1989. "Flore des Mascareignes. La Réunion, Maurice, Rodrigues".
- Böhle U. -R., Hilger H. H., Cerff R., Martin W. F. 1994. Non-coding chloroplast DNA for plant molecular systematics at the infrageneric level. In Desalle R., Wagner P., Schierwater B., Stoeit B. (eds.) "Molecular approaches to ecology and evolution", in press.
- Bremer B. 1991. Restriction data from chloroplast DNA for phylogenetic reconstruction : is there only one accurate way of scoring ? *Plant Systematics and Evolution*, 175 : 39-54.
- Bremer B., Jansen R. K. 1991. Comparative restriction site mapping of chloroplast DNA implies new phylogenetic relationships within *Rubiaceae*. *Am. J. Bot.*, 78 : 198-213.

## Références bibliographiques

- Bremer B., Struwe L.** 1992. Phylogeny of the *Rubiaceae* and the *Loganiaceae*: congruence or conflict between morphological and molecular data? *Am. J. Bot.*, 79 (10) : 1171-1184.
- Bremer K.** 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution*, 42 (4) : 795-803.
- Bridson D. M.** 1982. Studies in *Coffea* and *Psilanthus* (Rubiaceae) for the part two of "Flora of Tropical east Africa" : Rubiaceae. *Kew Bulletin*, 36 (4) : 817-860.
- Bridson D. M.** 1987. Nomenclatural notes on *Psilanthus*, including *Coffea* sect. *Paracoffea* (Rubiaceae tribe *Coffeae*). *Kew Bulletin*, 42 (2) : 453-460.
- Bridson D. M.** 1994. Additional note on *Coffea* (Rubiaceae) from tropical east Africa. *Kew Bulletin*, 49 (2) : 331-342.
- Bridson D. M., Verdcourt B.** 1988. "Flora of tropical East Africa - Rubiaceae (Part 2)", Polhill R. M. (eds.), 727 pp.
- Brown S. C., Bergounioux C., Tallet S., Marie D.** 1991. Flow cytometry of nuclei for ploidy and cell cycle analysis. In Negruțiu I., Gharti-Chetri G. (eds.) "A laboratory guide for cellular and molecular plant biology", Basel. pp. 326-345.
- Carvalho A., Medina Filho H. P., Fazuoli L. C., Guerreiro Filho O., Lima M. M. A.** 1991. Aspectos genéticos do cafeeiro. *Rev. Brasil. Genet.*, 14(1) : 135-183.
- Charrier A.** 1976. "La structure génétique des caféiers spontanés de la région malgache (*Mascarocoffea*). Leurs relations avec les caféiers d'origine africaine (*Eucoffea*). Thèse, Université d'ORSAY, 223 pp.
- Charrier A., Berthaud J.** 1975. Variation de la teneur en caféine dans le genre *Coffea*. *Café Cacao Thé*, XIX (n°4) : 251-264.
- Chase M. W., Soltis D. E., Olmstead R. G., Morgan D., Les D. H., Mishler B. D., Duvall M. R., Price R. A., Hills H. G., Qui Y. -L., Kron K., Rettig J. H., Conti E., Palmer J. D., Manhart J. R., Systma K. J., Michaels H. J., Kress W. J., Karol K. G., Clark W. D., Hedren M., Gaut B. S., Jansen R. K., Kim K. -J., Wimpee C. F., Smith J. F., Furnier G. R., Strauss S. H., Xiang Q. -Y., Plunkett G. M., Soltis P. S., Swensen S. M., Williams S. E., Gadek P. A., Quinn C. J., Eguiarte L. E., Golenberg E., Learn G. H., Graham S. W., Barrett S. C. H., Dayanandan S., Albert V. A.** 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcl*. *Annals of the Missouri Botanical garden*, 80 : 528-580.
- Chauvet M., Olivier L.** 1993. "La biodiversité enjeu planétaire", Paris. 415 pp.
- Chevalier A.** 1929. "Les caféiers du globe. Généralité sur les caféiers", Lechevalier P., 155 pp.
- Chevalier A.** 1942. "Les caféiers du globe. Iconographie des caféiers sauvages et cultivés et des Rubiacées prises pour des caféiers", Lechevalier P., 36 pp.
- Chevalier A.** 1947. Les caféiers du globe. Systématique des caféiers et faux caféiers. Maladies et insectes nuisibles. *Encycl. biol.*, XXVIII (Fas. III) : 1-356.
- Choo M. K., Soreng R. J., Davis J. I.** 1994. Phylogenetic relationships among *Puccinellia* and allied genera of Poaceae as inferred from chloroplast DNA restriction site variation. *Am. J. Bot.*, 81 (1) : 119-126.
- Clegg M. T.** 1990. Dating the Monocot-Dicot Divergence. *TREE*, vol. 5 n°1.
- Clegg M. T.** 1993. Chloroplast gene sequences and the study of plant evolution. *Proc. Natl. Acad. Sci. USA*, 90 : 363-367.

## Références bibliographiques

- Clegg M. T., Learn G. H., Golenberg E. M. 1990. Molecular evolution of chloroplast DNA. In Selander R. K., Clark A. J., Whittam T. S. (eds.) *"Evolution at the molecular level"*, Sunderland, pp. 135-149.
- Clegg M. T., Zurawski G. 1992. Chloroplast DNA and the study of plant phylogeny : present status and future prospects. In Soltis P. S., Soltis D. E., Doyle J. J. (eds.) *"Molecular systematics of plants"*, pp. 1-13.
- Clifford M. N., Gibson C. L., Rakotomalala J. J. R., Cros E., Charrier A. 1991. Caffeine from green beans of *mascarocoffea*. *Phytochemistry*, 30,12 : 4039-4040.
- Costich D. E., Ortiz R., Meagher T. R., Bruederle L. P., Vorsa N. 1993. Determination of ploidy level and nuclear DNA content in blueberry by flow cytometry. *Theor. Appl. Genet.*, 86 : 1001-1006.
- Cox A. V., Bennett M. D., Dyer T. A. 1992. Use of the polymerase chain reaction to detect spacer size heterogeneity in plant 5S-rRNA gene clusters and to locate such clusters in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.*, 83 : 684-690.
- Crawford D. J., Palmer J. D., Kobayashi M. 1990. Chloroplast DNA restriction site variation and the phylogeny of *Coreopsis* section *Coreopsis* (Asteraceae). *Am. J. Bot.*, 77 (4) : 552-558.
- Cros J., Combes M. -C., Chabrilange N., Duperray C., Monnot des Angles A., Hamon S. 1994. Nuclear DNA content in the sub-genus *Coffea* (Rubiaceae): inter- and intra specific variation in African species. *Can. J. Bot.*, à paraître.
- Dally A. M., Second G. 1990. Chloroplast DNA diversity in wild and cultivated species of rice (Genus *Oryza*, section *Oryza*). Cladistic-mutation and genetic-distance analysis. *Theor. Appl. Genet.*, 80 : 209-222.
- Darlu P., Tassy P. 1993. "Reconstruction phylogénétique. Concepts et méthodes", Masson, 229 pp.
- Demarly Y. 1975. Amélioration du Caféier liée aux progrès génétiques. *Ass. Sc. Inter. Café*, 7ème Colloque, Hambourg (9-14 juin 1975) : 423-435.
- Doebley J. 1992. Molecular systematics and crop evolution. In Soltis P. S., Soltis D. E., Doyle J. J. (eds.) *"Molecular systematics of plants"*, pp. 202-222.
- Doebley J., Bothmer R., Larson S. 1992. Chloroplast DNA variation and the phylogeny of *Hordeum* (Poaceae). *Am. J. Bot.*, 79 (5) : 576-584.
- Doebley J., Durbin M., Golenberg E. M., Clegg M. T., Ma D. P. 1990. Evolutionary analysis of the large subunit of carboxylase (*rbcL*) nucleotide sequence among the grasses (Gramineae). *Evolution*, 44 (4) : 1097-1108.
- Dolezel J. 1991. Flow cytometric analysis of nuclear DNA content in higher plants. *Phytochemical analysis*, 2 : 143-154.
- Dolezel J., Binarova P., Lucretti S. 1989. Analysis of nuclear DNA content in plant cells by flow cytometry. *Biologia Plantarum (Praha)*, 31(2) : 113-120.
- Dolezel J., Sgorbatti S., Lucretti S. 1992. Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiologia Plantarum*, 85 : 625-639.
- Dowling E., Moritz C., Palmer J. D. 1990. Restriction Site Analysis. In Hillis D. M., Moritz C. (eds.) *"Molecular Systematics"*, pp. 250-317.
- Doyle J. J. 1992. Gene Trees and Species Trees: Molecular Systematics as One-Character Taxonomy. *Systematic Botany*, 17 (1) : 144-163.
- Doyle J. J., Doyle J. L., Brown A. H. D. 1990. Chloroplast DNA polymorphism and phylogeny in the B genome of *Glycine* subgenus *Glycine* (Leguminosae). *Am. J. Bot.*, 77 (6) : 772-782.

## Références bibliographiques

- Duvall M. R., Clegg M. T., Chase M. W., Clark W. D., Kress W. J., Hills H. G., Eguiarte L. E., Smith J. F., Gaut B. S., Zimmer E. A., Learn G. H. 1993. Phylogenetic hypotheses for the monocotyledons constructed from *rbcL* sequence data. *Annals of the Missouri Botanical garden*, 80 : 607-619.
- Duvall M. R., Peterson P. M., Christensen A. H. 1994. Alliances of *Muhlenbergia* (Poaceae) within new world Eragrostideae are indentified by phylogenetic analysis of mapped restriction sites from plastid DNAs. *Am. J. Bot.*, 81 (5) : 622-629.
- Enomoto S. -I., Ogihara Y., Tsunewaki K. 1985. Studies on the origin of crop species by restriction endonuclease analysis of organellar DNA. Phylogenetic relationships among ten cereals revealed by the restriction fragment patterns of chloroplast DNA. *Jpn. J. Genet.*, 60 : 411-424.
- Faivre-Rampant P., Jeandroz S., Lefevre F., Lemoine M., Villar M., Berville A. 1992. Ribosomal DNA studies in poplars: *Populus deltoides*, *P. nigra*, *P. trichocarpa*, *P. maximowiczii*, and *P. alba*. *Genome*, 35 : 733-740.
- Farris J. S. 1970. Methods for computing Wagner trees. *Systematic Zoology*, 19 : 83-92.
- Farris J. S. 1989. The retention index and the rescaled consistency index. *Cladistics*, 5 : 417-419.
- Felsenstein J. 1989. PHYLIP-Phylogeny inference package (version 3. 2). *Cladistics*, 5 : 164-166.
- Felsenstein J. 1994. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39(4) : 783-791.
- Frascaria N., Maggia L., Michaud M., Bousquet J. 1993. The *rbcL* gene sequence from chestnut indicates a slow rate of evolution in the Fagaceae. *Genome*, 36 : 668-671.
- Freudenstein J. V., Doyle J. J. 1994. Plastid DNA, morphological variation, and the phylogenetic species concept: the *Corallorhiza maculata* (Orchidaceae) Complex. *Systematic Botany*, 19 (2) : 273-290.
- Gadek P. A., Quinn C. J. 1993. An analysis of relationships within the cupessaceae sensu stricto based on *rbcL* sequences. *Annals of the Missouri Botanical garden*, 80 : 581-586.
- Galbraith D. W., Harkins K. R., Madox J. M., Ayres N. M., Sharma D. P., Fiziozabaky E. 1983. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science*, 220 : 1049-1051.
- Gaut B. S., Clegg M. T. 1991. Molecular evolution of alcohol dehydrogenase 1 in members of the grass family. *Proc. Natl. Acad. Sci. USA*, 88 : 2060-2064.
- Gielly L. 1994. "ADN chloroplastique et phylogénies intragénériques", Thèse, Université Joseph Fourier - Grenoble I, 140 pp.
- Gielly L., Taberlet P. 1994. The use of chloroplast DNA to resolve plant phylogenies: non-coding versus *rbcL* sequences. *Mol. Biol. Evol.*, 11 (5) : 769-777.
- Gielly L., Taberlet P. 1994. Chloroplast DNA polymorphism at the intrageneric level and plant phylogenies. *Comptes rendus de l'Académie des Sciences, Paris*, 317 : 685-692.
- Gielly L., Taberlet P. 1994. A phylogeny of the European gentians inferred from chloroplast *trnL* (UAA) intron sequences. *Plant Systematics and Evolution*, (soumis).
- Glazmann J. -C., Nicolas D., Hamon S., Charrier A. 1993. La valorisation des ressources phytogénétiques tropicales. *La jaune et la rouge*, Novembre 93 : 8-9.
- Govindaraju D. R., Wagner D. B., Smith G. P., Dancik B. P. 1988. Chloroplast DNA variation within individual trees of a *Pinus banksiana*-*Pinus contorta* sympatric region. *Can. J. For. Res.*, 18 : 1347-1350.

## Références bibliographiques

- Graham M. J., Nickel C. D., Rayburn A. L. 1994. Relationship between genome size and maturity group in soybean. *Theor. Appl. Genet.*, 88 : 429-432.
- Guillaumet J. L., Halle F. 1978. Echantillonnage du matériel *coffea arabica* récolté en Ethiopie. *Bull. IFCC*, 14 : 13-18.
- Guillaumet J. L., Mangenot G. 1975. Aspects de la spéciation dans la flore malgache. *Boisseria*, 24 : 119-123.
- Gupta P. K., Tsuchia T. 1991. "Chromosome Engineering in Plants: Genetics, Breeding, Evolution. Part A", Elsevier (eds.), 627 pp.
- Hamby R. K., Zimmer E. A. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. In Soltis P. S., Soltis D. E., Doyle J. J. (eds.) "Molecular systematics of plants", pp. 50-91.
- Hamon P., Brizard J. P., Zoundjhekpon J., Duperray C., Borgel A. 1992. Etude des index d'ADN de huit espèces d'ignames (*Dioscorea* sp. ) par cytométrie en flux. *Can. J. Bot.*, 70 : 996-1000.
- Hamon S., Anthony F., Le Pierrès D. 1984. La variabilité génétique des caféiers de la section *Mozambicoffea* A. Chev. Précisions sur 2 espèces affines: *C. pseudozanguebariae* Bridson et *C. sp.* Bridson. *Bull. Mus. Hist. Nat. Andansonie*, 4 : 207-223.
- Harding K. 1991. Restriction enzyme mapping the ribosomal RNA genes in *Solanum tuberosum*: Potato rDNA restriction enzyme map. *Euphytica*, 54 : 245-250.
- Harlan J. R. 1969. Ethiopia: A center of Diversity. *Economic Botany*, 23 (X) : 309-314.
- Higgins D. G., Fuchs R., Blesby A. 1992. CLUSTAL: a new multiple sequence alignment program. *Comp. Appl. Biosci.*, 8 : 189-191.
- Hillis D. M., Huelsenbeck J. P. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. *Journal of Heredity*, 83(3) : 189-195.
- Hillis D. M., Larson A., Davis S. K., Zimmer E. A. 1990. Sequencing. In Hillis D. M., Moritz C. (eds.) "Molecular Systematics. ", pp. 318-370.
- Hiratsuka J., Shimada H., Whittier R., Ishibashi T., Sakamoto M., Masao M., Kondo C., Honji Y., Sun C. -R., Meng B. -Y., Li Y. -Q., Kanno A., Nishizawa Y., Hirai A., Shinozaki K., Sugiura M. 1989. The complete sequence of the rice (*Oriza sativa*) chloroplast genome : intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.*, 217 : 185-194.
- Hooglander N., Lumaret R., Bos M. 1993. Inter-intraspecific variation of chloroplast DNA of European *Plantago* spp. *Heredity*, 70 : 322-334.
- Hoot S. B., Reznicek A. A., Palmer J. D. 1994. Phylogenetic relationships in *Anemone* (Ranunculaceae) Based on morphology and chloroplast DNA. *Systematic Botany*, 19 (1) : 169-200.
- Hsiao C., Chatterton N. J., Asay K. H., Jensen K. B. 1994. Phylogenetic relationships of 10 grass species: an assessment of phylogenetic utility of the internal transcribed spacer region in nuclear ribosomal DNA in monocots. *Genome*, 37 : 112-120.
- Jansen R. K., Holsinger K. E., Michaels H. J., Palmer J. D. 1990. Phylogenetic analysis of chloroplast DNA restriction site data at higher taxonomic levels: an example from the Asteraceae. *Evolution*, 44(8) : 2089-2105.
- Jansen R. K., Michaels H. J., Palmer J. D. 1991. Phylogeny and Character Evolution in the Asteraceae Based on Chloroplast DNA Restriction Site Mapping. *Systematic Botany*, 16(1) : 98-115.

## Références bibliographiques

- Jansen R. K., Michaels H. J., Wallace R. S., Kim K. -J., Keeley S. C., Watson L. E., Palmer J. D. 1992. Chloroplast DNA variation in the Asteraceae : phylogenetic and evolutionary implications. In Soltis P. S., Soltis D. E., Doyle J. J. (eds.) *"Molecular systematics of plants"*, pp. 252-279.
- Jansen R. K., Palmer J. D. 1987. Chloroplast DNA from lettuce and *Barnadesia* (Asteraceae): structure, gene localization, and characterization of a large inversion. *Curr. Genet.*, 11 : 553-564.
- Johansson J. T., Jansen R. K. 1993. Chloroplast DNA variation and phylogeny of the *Ranunculeae*. *Pl. Syst. Evol.*, 187 : 29-49.
- Johnson L. A., Soltis D. E. 1994. *matK* DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. *Systematic Botany*, 19 (1) : 143-156.
- Kawagoe Y., Kikuta Y. 1991. Chloroplast DNA evolution in potato (*Solanum tuberosum* L.). *Theor. Appl. Genet.*, 81 : 13-20.
- Kenton A. 1984. Chromosome evolution in the *Gibasis linearis* group (Commelinaceae). III. DNA variation, chromosome evolution, and speciation in *G. venustula* and *G. heterophylla*. *Chromosoma*, 90 : 303-310.
- Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, 16 : 111-120.
- King L. M., Schaal B. A. 1989. Ribosomal DNA variation and distribution in *Rudbeckia Missourensis*. *Evolution*, 43 (5) : 1117-1119.
- Kluge A. G., Farris J. S. 1969. Quantitative phyletics and the evolution of anurans. *Systematic Zoology*, 18 (1) : 1-32.
- Kron K., Chase M. W. 1993. Systematics of the Ericaceae, Empetraceae, Epacridaceae and related taxa based upon *rbcL* sequence data. *Annals of the Missouri Botanical garden*, 80 : 735-741.
- Lanaud C., Tezenas du Montcel H., Jolivot M. P., Glaszmann J. C., Gonzalez de Leon D. 1991. Variation of ribosomal gene spacer length among wild and cultivated banana. *Heredity*, 68 : 118-127.
- Lashermes P., Combes M. C., Cros J. 1994. Use of non-radioactive digoxigenin-labelled DNA probes for RFLP analysis in coffee. Compte rendu du colloque "Techniques et utilisations des marqueurs moléculaire", Montpellier (France).
- Lashermes P., Cros J., Marmey P., Charrier A. 1993. Use of random amplified DNA markers to analyse genetic variability and relationships of *Coffea* species. *Genet. Res. Crop. Evol.*, 40 : 91-99.
- Lashermes P., Paillard M., Marmey P., Gavalda M. C., Couturon E., Petiard V., Charrier A. 1993. Toward the construction of a genetic map in coffee. *Ass. Sc. Inter. Café*, 15ème colloque, Montpellier (6-11 juin 1993) : 773-774.
- Leaver C. J. 1988. Le génome mitochondrial des végétaux supérieurs. *Biofutur*, janvier 1988 : 51-59.
- Leroy J. -F. 1961. Les faux Caféiers du genre *Argocoffea* (Pierre) Lebrun. *Op. cit.*, janvier-mars 1961 : 65-69.
- Leroy J. -F. 1967. Recherche sur les caféiers. Sur la classification biologique des Caféiers et sur l'origine et l'aire du genre *Coffea*. *Comptes rendus de l'Académie des Sciences, Paris*, 265 : 1043-1045.
- Leroy J. F. 1980. Evolution et taxogenèse chez les Caféiers (*Coffea* L., *Psilanthus* Hook. f. et *Nostolachma* Durand). Hypothèse sur leur origine. *Comptes rendus de l'Académie des Sciences, Paris*, 291 : 593-596.
- Leroy J. F. 1982. L'origine kenyane du genre *coffea* L. et la radiation des espèces à Madagascar. *Ass. Sc. Inter. Café*, 10ème colloque, Salvador-Bahia (11-14 octobre 1982) *Salvador* : 413-420.



## Références bibliographiques

- Livingstone D. A. 1982. Quaternary geography of Africa and the refuge theory. In Prance G. T. (eds.) "Biological diversification in the tropics", pp. 523-536.
- Louarn J. 1992. "La fertilité des hybrides interspécifiques et les relations génomiques entre caféiers diploïdes d'origine africaine (Genre *COFFEA* L. sous genre *COFFEA*).", Thèse, Université d'ORSAY, 178 pp.
- Louarn J. 1993. Structure génétique des caféiers africains diploïdes basée sur la fertilité des hybrides interspécifiques. *Ass. Sc. Inter. Café*, 15ème colloque, Montpellier (6-11 juin 1993) : 243-252.
- Maier H. G. 1993. Status of research in the field of non-volatile coffee components. *Ass. Sc. Inter. Café*, 15ème colloque, Montpellier (6-11 juin 1993) : 567-576.
- Manes Y. 1988. "Organisation du complexe multispécifique chez les caféiers: étude de la variabilité de 26 populations appartenant à l'espèce *Coffea brevipes*", ORSTOM (rapport VSNA), 35 pp.
- Maréchal-Drouard L., Weil J. H., Dietrich A. 1993. Transfer RNAs and transfer RNA genes in plants. *Ann. Rev. Plant Physiol.*, 44 : 13-32.
- Marie D., Brown S. C. 1993. A cytometric exercise in plant DNA histograms, with 2c values for 70 species. *Biol Cell*, 78 : 41-51.
- Martin W., Lydiat D., Brinkmann H., Forkmann G., Saedler H., Cerff R. 1993. Molecular Phylogenies in Angiosperm Evolution. *Mol. Biol. Evol.*, 10 (1) : 140-162.
- Mason R. J., Holsinger K. E., Jansen R. K. 1994. Biparental inheritance of the chloroplast genome in *Coreopsis* (Asteraceae). *Journal of Heredity*, 85(3) : 171-173.
- Masoud S. A., Johnson L. B., Sorensen E. L. 1990. High transmission of paternal plastid DNA in alfalfa plants demonstrated by restriction fragment polymorphic analysis. *Theor. Appl. Genet.*, 79 : 49-55.
- May C. E., Appels R. 1987. Variability and genetics of spacer DNA séquences between the ribosomal-RNA gene of hexaploid wheat (*Triticum aestivum*). *Theor. Appl. Genet.*, 74 : 617-624.
- Mayer M. S., Soltis P. S. 1994. Chloroplast DNA phylogeny of *Lens* (Leguminosae): origin and diversity of the cultivated lentil. *Theor. Appl. Genet.*, 87 : 773-781.
- Michaels H. J., Scott K. M., Olmstead R. G., Szaro T., Jansen R. K., Palmer J. D. 1993. Interfamilial relationships of the Asteraceae: insights from *rbcL* sequence variation. *Annals of the Missouri Botanical garden*, 80 : 742-751.
- Michaelson M. J., Price H. J., Ellison J. R., Johnston J. S. 1991. Comparison of plant DNA contents determined by feulgen microspectrophotometry and laser flow cytometry. *Am. J. Bot.*, 78 (2) : 183-188.
- Michaelson M. J., Price H. J., Johnson J. R., Ellison J. R. 1991. Variation of nuclear DNA content in *Helianthus annuus* (Asteraceae). *Am. J. Bot.*, 78 : 1238-1243.
- Natali L., Cavallini A., Cionini G., Sassoli O., Cionini P. G., Durante M. 1993. Nuclear DNA changes within *Helianthus annuus* L. : changes within single progenies and their relationships with plant development. *Theor. Appl. Genet.*, 85 : 506-512.
- Namur (de) C. 1990. Aperçu sur la végétation de l'Afrique centrale atlantique. In ORSTOM (eds.) "Paysage quaternaire de l'Afrique centrale atlantique", pp. 60-67.
- Nee S., Holmes E. C., May R. M., Harvey P. H. 1994. Extinction rates can be estimated from molecular phylogenies. *Phil. Trans. R. Soc. Lond.*, 344 : 77-82.
- Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S., Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H., Ozeki H. 1986. Complete nucleotide sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Plant Molecular Biology Reporter*, 4 : 148-175.

## Références bibliographiques

- Olmstead R. G., Bremer B., Scott K. M., Palmer J. D. 1993. A parsimony analysis of the Asteridae sensu lato based on *rbcl* sequences. *Annals of the Missouri Botanical garden*, 80 : 700-722.
- Olmstead R. G., Palmer J. D. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *Am. J. Bot.*, 81(9) : 1205-1224.
- Orozco-Castillo C., Chalmers K. J., Powell W. 1994. Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. *Theor. Appl. Genet.*, 87 : 934-940.
- Palmer J. D. 1985. Comparative organization of chloroplast genomes. *Ann. Rev. Genet.*, 19 : 325-354.
- Palmer J. D. 1987. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *American Naturalist*, 130 : S6-S29.
- Palmer J. D. 1992. Mitochondrial DNA in plant systematics: applications and limitations. In Soltis P. S., Soltis D. E., Doyle J. J. (eds.) *"Molecular systematics of plants"*, pp. 36-49.
- Palmer J. D., Jansen R. K., Michaels H. J., Chase M. W., Manhart J. R. 1988. Chloroplast DNA variation and plant phylogeny. *Annals of the Missouri Botanical garden*, 75 : 1180-1206.
- Pocs T. 1975. Affinities between the bryoflora of East Africa and Madagascar. *Boisseria*, 24 : 125-128.
- Pradhan A. K., Prakash S., Mukhopadhyay A., Pental D. 1992. Phylogeny of *Brassica* and allied genera based on variation in chloroplast and mitochondrial DNA patterns: molecular and taxonomic classifications are incongruous. *Theor. Appl. Genet.*, 85 : 331-340.
- Price H. J. 1988. DNA content variation among higher plants. *Annals of the Missouri Botanical garden*, 75 : 1248-1257.
- Price R. A., Palmer J. D. 1993. Phylogenetic relationships of the Geraniaceae and Geraniales from *rbcl* séquence comparisons. *Annals of the Missouri Botanical garden*, 80 : 661-671.
- Qiu Y. -L., Chase M. W., Les D. H., Parks C. R. 1993. Molecular phylogenetics of the Magnoliidae: cladistic analyses of nucléotide séquences of the plastid gene *rbcl*. *Annals of the Missouri Botanical garden*, 80 : 587-606.
- Rakotomalala J. J. R. 1992. "Diversité biochimique des caféiers - Particularités des caféiers sauvages de la région malgache". Thèse, Université Montpellier II, pp. 219.
- Rakotomalala J. J. R., Cros E., Charrier A. 1993. Quelques acides phénols marticuliers des fèves de caféiers sauvages malgaches et africains. *Ass. Sc. Inter. Café*, 15ème colloque, Montpellier (6-11 juin 1993) : 637-643.
- Rakotomalala J. J. R., Cros E., Charrier A., Anthony F., Noirot M. 1993. Marqueurs biochimiques de la diversité des caféiers. *Ass. Sc. Inter. Café, Montpellier*, 15 : 47-55.
- Raven P. H., Axelrod D. I. 1974. Angiosperm biogeography and past continental movements. *Annals of the Missouri Botanical garden*, 61(3) : 539-673.
- Reboud X., Zeyl C. 1994. Organelle inheritance in plants. *Heredity*, 72 : 132-140.
- Rieseberg L. H., Soltis D. E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants*, 5(1) : 65-84.
- Rose R., Johnson L. B., Kemble R. J. 1986. Restriction endonuclease studies on the chloroplast and mitochondrial DNAs of alfalfa (*Medicago sativa* L. ) protoclonés. *Plant Molecular Biology*, 6 : 331-338.

## Références bibliographiques

- Saghai-Maroo M. A., Soliman K. M., Jorgensen R. A., Allard R. W. 1984. Ribosomal DNA spacer-length polymorphism in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. USA*, 81 : 8014-8018.
- Sakata Y., Nishio T., Matthews P. J. 1991. Chloroplast DNA analysis of eggplant (*Solanum melongena*) and related species for their taxonomic affinity. *Euphytica*, 55 : 21-26.
- Sambrook J., Fritsch E. F., Maniatis T. 1989. "Molecular cloning. A laboratory manual", Cold Spring Harbor Laboratory Press.
- Sano Y., Sano R. 1990. Variation of the intergenic spacer region of ribosomal DNA in cultivated and wild rice species. *Genome*, 33 : 209-218.
- Santa Ram A., Sreenivasan M. S., Vishveshwara S. 1982. Chemotaxonomy of Mozambicoffea. *J. Coffee Res.*, 12, 2 : 42-46.
- Savolainen V., Manen J. F., Douzery E., Spichiger R. 1994. Molecular Phylogeny of Families Related to Celastrales based on rbcL 5' Flanking Sequences. *Molecular Phylogenetics and Evolution*, 3 (1) : 27-37.
- Schaal B. A., Learn G. H. 1988. Ribosomal DNA variation within and among plant populations. *Annals of the Missouri Botanical garden*, 75 : 1207-1216.
- Schilling E. E., Panero J. L., Eliasson U. H. 1994. Evidence from chloroplast DNA restriction site analysis on the relationships of *Scalesia* (Asteraceae: Heliantheae). *Am. J. Bot.*, 81 (2) : 248-254.
- Schmit V., Jardin (du) P., Baudoin J. P., Debouck D. G. 1993. Use of chloroplast DNA polymorphism for the phylogenetic study of seven *Phaseolus* taxa including *P. vulgaris* and *P. coccineus*. *Theor. Appl. Genet.*, 87 : 506-516.
- Scott N. S., Timmis J. N., Ayliffe M. A. 1991. Chloroplast genes in higher plants. *Aust. J. Plant Physiol.*, 18 : 533-541.
- Shield R. 1993. Pastoral synteny. *Nature*, 365 : 297-298.
- Shinozaki K., Ohme M., Tanaka M., Wakasugi T., Hayashida N., Matsubayashi T., Zaita N., Chunwongse J., Obokata J., Yamaguchi-Sinozaki K., Ohto C., Torazawa K., Meng B. Y., Sugita M., Deno H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A., Tohdoh N., Shimada H., Sugiura M. 1986. The complete nucleotide sequence of the tobacco chloroplast genome. *Plant Molecular Biology Reporter*, 4 : 110-147.
- Silwar R., Lüllmann C. 1993. The aroma composition of the coffee beverage. Quantitative determination of steam-volatile aroma constituents. *Ass. Sc. Inter. Café*, 15ème colloque, Montpellier (6-11 juin 1993) : 873-879.
- Sim L., Price H. J. 1985. Nuclear DNA content variation in *Helianthus* (Asteraceae). *Am. J. Bot.*, 72 : 1213-1219.
- Smith A. B. 1994. Rooting molecular trees: problems and strategies. *Biological Journal of the Linnean Society*, 51 : 279-292.
- Sobral B. W. S., Braga D. P. V., LaHood E. S., Keim P. 1994. Phylogenetic analysis of chloroplast restriction enzyme site mutation in the *Saccharinae* Griseb. subtribe of the *Andropogoneae* Dumort. tribe. *Theor. Appl. Genet.*, 87 : 843-853.
- Soltis D. E., Morgan D. R., Grable A., Soltis P. S., Kuzoff R. 1993. Molecular systematics of Saxifragaceae sensus stricto. *Am. J. Bot.*, 80 (9) : 1056-1081.
- Speer K., Sehat N., Montag A. 1993. Fatty acids in coffee. *Ass. Sc. Inter. Café*, 15ème colloque, Montpellier (6-11 juin 1993) : 83-592.

## Références bibliographiques

- Spichiger R., Savolainen V., Manen J. F. 1993. Systematic affinities of Aquifoliaceae and Icacinaceae from molecular data analysis. *CANDOLLEA*, 48 : 459-464.
- Steele K. P., Vilgalys R. 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *matK*. *Systematic Botany*, 19 (1) : 126-142.
- Sugiura M. 1992. The chloroplast genome. *Plant Molecular Biology*, 19 : 149-168.
- Suh Y., Thien L. B., Reeve H. E., Zimmer E. A. 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. *Am. J. Bot.*, 80(9) : 1042-1055.
- Swensen S. M., Mullin B. C. 1994. Phylogenetic affinities of Datisceae based on an analysis of nucleotide sequences from the plastid *rbcl* gene. *Systematic Botany*, 19(1) : 157-168.
- Systma K. J., Gottlieb L. D. 1986. Chloroplast DNA evolution and phylogenetic relationships in *Clarkia* sect. *Peripetasma* (Onagraceae). *Evolution*, 40(6) : 1248-1261.
- Taberlet P., Gielly L., Pautou G., Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, 17 : 1105-1109.
- Terauchi R., Terachi T., Tsunewaki K. 1991. Intraspecific variation of chloroplast DNA in *Dioscorea bulbifera* L. *Theor. Appl. Genet.*, 81 : 461-470.
- Van Ham R. C. H. J., Hart H. 1994. Evolution of *Sedum* series *Rupestria*: evidence from chloroplast DNA and biosystematic data. *Plant Systematics and Evolution*, 190 : 1-20.
- Vishveshwara S., Ahmed J. 1975. Variability in Coffee. II. Studies on flower number in species. *Indian coffee*, XXXIX : 268-273.
- Wagner D. B., Furnier G. R., Saghai-Marooof M. A., Williams S. M., Dancik B. P., Allard R. W. 1987. Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. *Proc. Natl. Acad. Sci. USA*, 84 : 2097-2100.
- Wakamiya I., Newton R. J., Johnston J. S., Price H. J. 1993. Genome size and environmental factor in the genus *Pinus*. *Am. J. Bot.*, 80 (11) : 1235-1241.
- Weir B. S., Swofford D. L., Olsen G. J., Hillis D. M., Moritz C. 1990. Analysis. In Hillis D. M., Moritz C. (eds.) *"Molecular Systematics"*, pp. 373-515.
- White F. 1979. *"The Guineo-congolian region and its relationships to other phytocoria"*, Bull. Jard. Bot. Nat. Belg. (eds.), 55 pp.
- Wilson H. D., Doebley J., Duvall M. 1992. Chloroplast DNA diversity among wild and cultivated members of *Cucurbita* (Cucurbitaceae). *Theor. Appl. Genet.*, 84 : 859-865.
- Wolfe K. H., Li W. -H., Sharp P. M. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA*, 84 : 9054-9058.
- Wolff K., Schaal B. 1992. Chloroplast DNA variation within and among five *Plantago* species. *J. evol. biol.*, 5 : 325-344.
- Xiang Q. -Y., Soltis D. E., Morgan D. R., Soltis P. S. 1993. Phylogenetic relationships of *Cornus* L. sensus lato and putative relatives inferred from *rbcl* sequence data. *Annals of the Missouri Botanical garden*, 80 : 723-734.
- Zurawski G., Clegg M. T. 1987. Evolution of higher-plant chloroplast DNA-encoded genes: implications for structure-function and phylogenetic studies. *Ann. Rev. Plant Physiol.*, 38 : 391-418.

## ANNEXES

## ANNEXE 1

**Extraction d'ADN total**

- Le matériel végétal (feuilles) est lyophilisé puis broyé en poudre fine à l'aide d'un broyeur à hélice (CYCLOTECH). Les poudres sont conservées en pots étanches à -20°C.
- Le matériel végétal (sous forme de poudre) est transféré dans un tube. Lorsque la poudre est totalement décongelée, on ajoute 15 ml de tampon d'extraction préchauffé à 65°C. Après homogénéisation de la suspension, le mélange est incubé durant 1h 30 à 65°C sous agitation douce.
- Après cette incubation, laisser le mélange revenir à température ambiante (environ 10 minutes).
- Procéder à une première extraction : ajouter 7,5 ml de chloroforme/isoamylalcool 24/1, incuber 10 minutes à température ambiante sous agitation, puis centrifuger 10 minutes à 1500g et toujours à température ambiante.
- Récupérer la phase supérieure et procéder à une seconde extraction.
- La phase supérieure est incubée durant 30 minutes en présence de 350 µg de RNase pancréatique.
- La précipitation est effectuée tube après tube, par addition de 15 ml d'isopropanol à température ambiante. Dès que la pelote se forme, elle est récupérée à l'aide d'une pipette pasteur courbée en hameçon et placée dans un tube contenant 1 ml de TE jusqu'à complète reprise (une nuit).
- Une seconde précipitation est réalisée à l'aide de 2,5 ml d'éthanol et de 50 µl de NaCl 5M. La pelote est récupérée à l'aide d'une pipette pasteur courbée en hameçon et placée durant 20 minutes dans un tube contenant 3 ml de tampon de lavage.
- La pelote est ensuite soigneusement égouttée, puis transférée dans un tube contenant 500 µl de TE.

**Tampon d'extraction**

Tris HCl 0,1M pH 8

NaCl 1,25M

EDTA 0,02 M

MATAB 2% (Mixed Alkyl Trimethyl Ammonium Bromide)

β-mercaptoethanol 0,1%

**Tampon de lavage**

Ethanol pur 76 %

Acétate de Sodium 0,2M

**TE**

Tris HCl 0,1M pH 8

EDTA 10mM pH 8

## ANNEXE 2

**Préparation du tampon de dépôt contenant des standards internes de poids moléculaire**

Le tampon de dépôt est préparé de façon à avoir une bande (standard interne) à 24,8 kb et une autre à 1,5 kb, chacune correspondant à 150 pg d'ADN. De cette façon les bandes de standard interne sont invisibles sur le gel coloré au Bet mais sont révélées après hybridation avec de l'ADN total de bactériophage  $\lambda$  marqué au  $^{32}\text{P}$  (que l'on mélange à la sonde spécifique). Ces bandes permettent de visualiser d'éventuelles distorsions en cours de migration.

**Isolement préalable des fragments**

- Deux lots d'ADN de bactériophage  $\lambda$  (Boeringer) sont digérés parallèlement par XbaI et Hind III (Boeringer).
- Après migration sur gel d'agarose à bas point de fusion (low melting), les bandes à 24800 pb et 1500 pb issues respectivement des digestions par XbaI et Hind III sont isolées et purifiées (protocole décrit dans Maniatis).

**Clonage, amplification, et purification des fragments**

- Les fragments sont clonés dans pUC 18 aux sites correspondants (XbaI et Hind III), selon un protocole standard (protocole décrit dans Maniatis). Des bactéries DH5 $\alpha$  rendues compétentes, sont transformées par le pUC 18 modifié (protocole décrit dans Maniatis). La sélection des clones transformés par les plasmides modifiés est effectué sur milieu LB solide contenant de l'Ampicilline (100 mg/ml), de l'IPTG 30 mM et XGAL 0,12% (colonies blanches).
- Des stocks sont effectués pour chacun des deux clones, dans un mélange de LB liquide et de glycérol (75/25), et conservés en aliquotes de 100  $\mu\text{l}$  à -80°C.
- Parallèlement les 2 clones sont amplifiés, leur ADN plasmidique extrait, les inserts isolés. Des stocks de solution d'ADN concentré des deux fragments sont constitués.

**Tampon de dépôt (3  $\mu\text{l}$  par échantillon à déposer):**

Bleu de Bromophénol	0,25%
Glycérol	30%
EDTA pH 8	0,1M

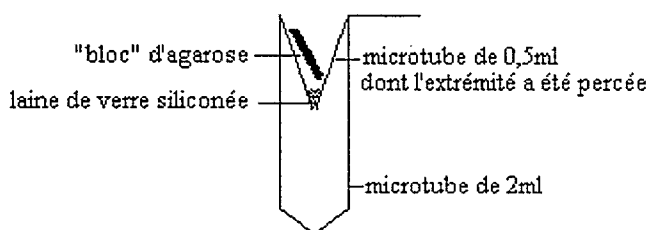
ADN fragment 24,8 kb 50 ng/ml

ADN fragment 1,5 kb 50 ng/ml

## ANNEXE 3

**Purification par cycles de congélation/centrifugation**

Le fragment à purifier est isolé après migration sur gel d'agarose : La bande d'agarose contenant le fragment à purifier (visualisée sous lumière U.V. après coloration au Bet) est découpée au scalpel et placée dans un microtube comme décrit ci dessous:



L'ensemble est placé à  $-20^{\circ}\text{C}$  pour 30 minutes, puis centrifugé 15 minutes à 15000g. 50 à 200  $\mu\text{l}$  de TE (en fonction du volume du bloc d'agarose) sont rajoutés au dessus de l'agarose.

L'ensemble est à nouveau placé à  $-20^{\circ}\text{C}$  puis centrifugé et recouvert de TE. Ce cycle est effectué au moins trois fois.

Le tube supérieur (contenant les "restes" d'agarose) est éliminé.

Le contenu du tube inférieur est précipité par addition de 3 volumes d'éthanol et de NaCl à une concentration finale de 80 mM. Après précipitation et séchage, le culot est repris dans un volume de tampon (Tris HCl ou TE) adéquat.

Ce protocole a été utilisé pour l'isolement de fragments de 50 à plus de 5000 pb. Dans la majorité des cas, le rendement est compris entre 50 et 80%.



ANNEXE 4

---

**Pré-hybridation, hybridation, lavage de stringence**

Toutes ces opérations sont réalisées à 65°C, dans un four à hybridation (Appligène). Les étapes de pré-hybridation et d'hybridation se font en tube (une membrane par tube). Les lavages de stringence sont faits sous agitation dans un bac avec un gros volume de tampon (plusieurs membranes par bac)

**Pré-hybridation**

Disposer une membrane par tube, et ajouter 20 ml de 2X SSC. Placer les tubes dans le support du four, mettre en rotation et préchauffer à 65°C. Après 10 min. vérifier que la membrane adhère aux parois internes du tube (sans bulles). Eliminer le 2X SSC et ajouter 20 ml de solution de pré-hybridation préalablement préchauffée à 65°C plus 50 µl de solution d'ADN de sperme de saumon dénaturé. Remettre les tubes en rotation dans le four pour au moins 4 heures.

**Préparation de la sonde marquée**

Le marquage est effectué selon le protocole du kit "random priming" (Amersham) légèrement modifié : pour 10 ng d'ADN à marquer (sonde ou ADN de bactériophage λ) dans un volume de 50 µl, on utilise 15 µCurie d'αdCT<sup>32</sup>P. La réaction s'effectue à 37°C durant 90 min. A la fin de la réaction, le mélange est simplement dénaturé par ébullition 5' puis plongé dans la glace. Les 50 µl de sonde marquée, plus 5 µl de la solution d'ADN de bactériophage λ marqué sont ensuite transférés dans un tube contenant 20 ml de tampon d'hybridation préchauffé à 65°C

**Hybridation**

Vider la solution de pré-hybridation et la remplacer par la solution d'hybridation contenant la sonde et le λ marqués. Replacer les tubes dans le four et mettre en rotation pour une nuit.

## ANNEXE 4 (suite)

### Lavage de stringence

Vider le contenu du tube d'hybridation (dans un conteneur de récupération de solution contenant du  $^{32}\text{P}$ ). Faire un premier lavage 5 min. avec 20 ml de 2X SSC, suivi de plusieurs lavages, en alternant le sens de rotation des tubes:

- 2X SSC 0,1% SDS (15 min.)
- 1X SSC 0,1% SDS (15 min.)
- 0,5X SSC 0,1% SDS (15 min.)
- 0,2X SSC 0,1% SDS (15 min.)

Mettre les membranes en cassette entre deux feuilles de film plastique laisser exposer avec un film autoradiographique pour 4 heures ou plus en fonction de l'intensité du signal.

### Déhybridation

placer les membranes dans un bac sous agitation, les recouvrir de SDS 1% bouillant (3 fois 20 min.).

#### Dehnart 50X

Ficoll 1%

1% PVP 1%

Albumine de Sérum Bovin (BSA) 1%

dans H<sub>2</sub>O milliQ.

la solution est filtrée sur 45 µm,

puis stockée à -20°C

#### 20X SSC

protocole : Sambrook & *al.* (1989).

#### Solution d'ADN de sperme de saumon

protocole Sambrook & *al.* (1989),

avant utilisation dénaturer 5 min. à 98°C

refroidir rapidement dans la glace

#### Solution de pré-hybridation

SSC 3X

solution de Dehnart 5X

SDS 0,1%

PEG 8000 6%

dans H<sub>2</sub>O milliQ stérile.

#### Solution d'hybridation

SSC 3X

solution de Dehnart 2X

SDS 0,1%

PEG 8000 6%

dans H<sub>2</sub>O milliQ stérile.

## ANNEXE 5

**Gels d'électrophorèse****gels d'agarose:**

Plusieurs types d'agaroses sont utilisés

"normal" dans la plupart des utilisations (GIBCO)

Low melting pour l'isolement de fragments (GIBCO)

Metaphor pour améliorer la résolution entre petits fragments (FMC BioProducts)

L'agarose est utilisé à des concentrations variables

0,8 à 1% en routine (dosage d'ADN, isolement de sondes, migrations pour RFLP...)

2% lorsque les fragments attendus ont des poids moléculaires plus faibles (RAPD, petites sondes)

3% avec du Metaphor pour une très bonne résolution sur les petits fragments (digestion de fragments PCR)

La quantité d'agarose nécessaire est pesée, puis mélangée au TBE 1X. La solution est chauffée jusqu'à dissolution complète de l'agarose, puis après refroidissement partiel coulée dans le moule adéquat, le peigne est positionné.

La migration s'effectue sur des gels horizontaux immergés dans du TBE 1X à voltage constant (30 à 120 volts suivant le type de migration).

**Gels de polyacrylamide:**

Nettoyer les plaques de verre et les "espaceurs" à l'alcool. Monter le système.

Préparer le mélange extemporanément:

	gel à 5%	gel à 10%
sol. d'acrylamide à 40%	6,25 ml	12,5 ml
TBE 10X	5 ml	5 ml
sol. de persulfate d'ammonium à 10%	350 µl	350 µl
H2O bidistillée qsp 50 ml	38,36 ml	32,11 ml
TEMED	35 µl	35 µl

Couler le gel entre les deux plaques de verre verticales en évitant les bulles. Positionner le peigne.

Après polymérisation (environ 30 minutes) retirer le peigne et nettoyer les dents avec du TBE 1X.

Placer le système dans la cuve. La migration s'effectue en présence de TBE 1X, à 25 mA soit 60 à 70 volt.

ANNEXE 5 (suite)

**Gels de polyacrylamide pour le séquençage:**

Nettoyer très soigneusement les plaques de verre et les "espaceurs" (0,4 mm), terminer par un rinçage à l'eau bidistillée. Préparer le montage plaques+espaceurs.

pour un gel de 100 ml mélanger,

50g d'urée

15 ml de solution d'acrylamide à 40%

29 ml d'H<sub>2</sub>O bidistillée

1g de résine amberlite

Chauffer légèrement et agiter jusqu'à dissolution complète de l'urée. Filtrer le mélange sous vide sur un filtre de porosité 0,2µm.

Ajouter 10 ml de TBE 10X à la solution filtrée puis ajuster le volume à 100 ml avec de l'eau bidistillée.

Dégazer la solution.

Ajouter 500 µl d'une solution de persulfate d'ammonium et 56 µl de TEMED.

Couler le gel entre les deux plaques de verre, à l'horizontale. Disposer le peigne. Attendre 2 heures pour s'assurer que la polymérisation est complète.

**TBE 10X**

107,8g Tris

55,0g Acide borique

8,2g EDTA

qsp 1l H<sub>2</sub>O bidistillée

Autoclaver

**Solution d'acrylamide à 40%**

38g Acrylamide

2g Bis-Acrylamide

qsp 100 ml H<sub>2</sub>O bidistillée

désioniser et filtrer

## ANNEXE 6

## Amplifications par PCR

Elles sont réalisées sur un trioblock BIOMETRA à effet Peltier, utilisant des microtubes de 0,5 ml. Les pentes sont fixées à 1°C par seconde.

### Amplifications spécifiques:

Les amplifications sont réalisées dans un volume de 25, 50 ou 100µl, en fonction de la quantité de produit (bande amplifiée) désirée.

Amorces	0,5µM de chaque
dNTP	100µM (25µM de chacun)
AmpliTaq DNA polymérase	1 unité pour 25 ou 50 µl, 2 unités pour 100.
Tampon spécifique	1X
ADN total	50 à 200 ng

cycles utilisés: 5 minutes à 95°C  
 1 minute à 95°C I  
 1 minute à **température d'hybridation (variable)** I X 35 cycles  
**temps variable** à 72°C I  
**temps variable** à 72°C (élongation terminale)  
 conservation à 4°C

### amplifications de type RAPD:

Les amplifications sont réalisées dans un volume réactionnel de 25µl.

Amorce	0,4µM
dNTP	600µM (150mM de chacun)
AmpliTaq DNA polymérase	2 unités
Tampon spécifique	1X
ADN total	50ng

Avant amplification le mélange réactionnel est recouvert de 3 gouttes de paraffine liquide.

cycle utilisé: 2 minutes à 95°C  
 1 minute à 95°C I  
 1 minute à 34°C I X 45 cycles  
 2 minutes à 72°C I  
 7 minutes à 72°C  
 conservation à 4°C avant dépôt

## ANNEXE 7

**Séquençage de produits PCR** (Kit dideoxitermineur d'Applied Biosystem-Perkin Elmer)

Le fragment à séquencer est amplifié au cours d'une première étape d'amplification spécifique (voir annexe 5), puis isolé et purifié selon le protocole décrit en annexe 3. Après purification l'ADN est repris dans du Tris HCl 5mM à pH 8, et est soigneusement dosé.

## préparation du Prémix:

5X TACS buffer	4 µl /échantillon
dNTP Mix	1 µl /échantillon
Dye dideoxy A termineur	1 µl /échantillon
Dye dideoxy C termineur	1 µl /échantillon
Dye dideoxy G termineur	1 µl /échantillon
Dye dideoxy T termineur	1 µl /échantillon
Amplitaq DNA polymérase	5 µl /échantillon

La réaction de séquence est effectuée dans un volume final de 20 µl:

Prémix	9,5µl
fragment d'ADN purifié	100 ng (préalablement dénaturé 10 minutes à 95°C, et 2 minutes à 0°C)
Amorce	3,75 pM (attention, une seule!)
H2O	qsp 20 µl

l'ensemble du mélange est recouvert de 3 gouttes de paraffine liquide avant amplification.

cycle d'amplification utilisé	1 minute 30 secondes à 96°C
	30 secondes à 96°C    ]
	15 secondes à 50°C    ] X 25 cycles
	4 minutes à 60°C       ]
	conservation à 4°C

Après l'amplification, on récupère la phase aqueuse (inférieure) à laquelle on ajoute 80 µl d'eau stérile.

On réalise ensuite deux extractions successives au phénol/H<sub>2</sub>O/chloroforme, à température ambiante.

L'ADN contenu dans la phase supérieure est finalement précipité par addition de 15 µl d'acétate de sodium 2M à pH 4,5 et 300 µl d'éthanol pur. Après centrifugation et lavage à l'éthanol 70%, le culot est séché par évaporation sous vide (speed vac). Juste avant le dépôt, les échantillons sont repris dans 4 µl de tampon de dépôt, et l'ensemble est dénaturé 2 minutes à 93°C.

Après vérification de sa qualité (absence de bulles, polymérisation homogène) le gel de séquence préparé au préalable est placé dans le séquenceur (Modèle 373A Applied Biosystems). Le système est mis en "pré-run" pour environ 30 minutes, jusqu'à égalisation des seuils de fluorescence. La migration, qui s'effectue durant 12 heures à 1200Volts.

## ANNEXE 8

## Isolement des noyaux pour la quantification du contenu en ADN nucléaire par cytofluorométrie en flux.

### Purification de noyaux (pour la coloration Iodure de Propidium)

1,5 g de feuilles fraîches sont broyées en présence d'azote liquide, puis mélangés avec 10 ml de tampon A. Le mélange est filtré sur une membrane de porosité 50  $\mu\text{m}$  afin d'éliminer les plus gros débris. La solution est ensuite centrifugée à 3000g à 4°C durant 15 min., le culot est resuspendu dans 10 ml de tampon B, avant d'être à nouveau centrifugé (3000g à 4°C durant 15 min.). Le culot est délicatement mélangé avec 5 ml de tampon B. Ce mélange est déposé dans un tube contenant 20 ml de tampon C. L'ensemble est centrifugé à 16000g et à 4°C durant 45 min. Ce dernier culot est stocké à - 80°C avant analyse.

Avant le passage en cytofluorométrie, le culot décongelé est repris dans 500  $\mu\text{l}$  de tampon PBS à pH 7,4, puis incubé durant 5 min. en présence d'iodure de propidium à la concentration finale de 160  $\mu\text{g/ml}$ .

### Isolement de noyaux (pour la coloration DAPI)

Le protocole est adapté à partir du protocole "classique" d'isolement des noyaux par lacération de feuilles fraîches (Galbraith *et al.*, 1983).

500 mg de feuilles fraîches sont lacérées dans une boîte de pétrie en présence de 1 ml de tampon commercial CHEMUNEX (référence R09-220-500 CHEMUNEX, Maisons-Alfort, France). Le mélange est incubé durant 20 minutes à température ambiante, à l'obscurité en présence de DAPI à la concentration finale de 10  $\mu\text{g/ml}$ . Avant analyse le mélange est filtré sur une membrane de porosité 17  $\mu\text{m}$ .

#### Tampon A

0,4M sucrose  
0,05M Tris  
2mM  $\text{CaCl}_2$   
0.4%  $\beta$ -mercaptoethanol

#### Tampon C

2M sucrose  
0,05M Tris  
2mM  $\text{CaCl}_2$

#### Tampon B

0,25 M sucrose  
0,05M Tris  
2mM  $\text{CaCl}_2$

#### PBS pH 7,4

$7.74 \cdot 10^{-2}\text{M Na}_2\text{HPO}_4$   
 $2.26 \cdot 10^{-2}\text{M NaH}_2\text{PO}_4$



## Thèses et documents microfichés

### Mots-clés

ADN chloroplastique, Phylogénie, Séquence intergénique, Afrique, *Coffea*, RFLP, Taille du génome.



209-213, rue La Fayette  
75480 Paris cedex 10

ISSN : 0767-922X  
ISBN : 2-7099-1313-5

Diffusion :  
32, avenue Henri Varagnat  
93143 Bondy cedex